Quantitative Proteomic Studies Of Protein Misfolding & Disease

Adam Cryar, Mchem (Hons)

University College London

Institute of Structural & Molecular Biology

Thesis submitted for the degree of Doctor of Philosophy

September 2015

CONTENTS

CHAPTERS

Declaration

I, Adam Cryar, declare that the work presented in this thesis is my own. Where information has been derived from other sources, or is of a collaborative nature, I confirm that this has been explicitly indicated in the text.

Acknowledgments

It takes a thousand voices to tell a single story

Native American Proverb

This thesis is the result of the lasting support and friendship of the many people in my life, past and present. Although there are too many to name in person I am pleased to have the chance to acknowledge those closest and most important to me.

I would first like to thank my supervisor Kostas Thalassinos for his support, advice and continued enthusiasm throughout my PhD. In particular Kostas' supervision has provided me with the freedom to explore my ideas, improve my weaknesses and to develop both my scientific and life skills. Not only that, he has given me invaluable career advice, home-buying advice and most importantly DIY tips!

My entire thesis has been the result of excellent collaborations and all need to be thanked. Firstly I would like to thank David for his countless hours of work, advice and proof reading. Never have I met a man with such a huge amount of knowledge, about seemingly everything, but with a particular speciality in German proverbs. Without your help this all would have been far more difficult. I would also like to thank Bibek Gooptu and Peg Nyon for their hard work, biological knowledge and insight. This side collaboration turned out to be a most fruitful chapter and I am grateful for all of your input that helped produce such a successful collaboration. Finally I would like to thank Fiona Kerr for all of her help. For your patience in teaching me the fly lab skills, for her work during the collaboration and for all the scientific advice you gave in the succeeding years.

I would obviously like to thank my lab colleagues, past and present who have helped make my four years at UCL enjoyable. I would like to thank Rich, Jun, Harpal and Ganesh for making me so welcome when I first joined. Ganesh in particular was a great mate to have in office and at lunchtimes for venting frustrations at. Thanks to Zoja for good scientific discussions and a great trip to America. I would also like to thank Kieran for many interesting science and non-science-related chats in the office.

I would like to thank all of my friends. Those who have helped me maintain my perspective on life outside of this thesis and those who were a part of me getting here. These include Matt T, Tom H, Steve T, Sam H, Patterson and Laura & Rob (the Hammersmith years)

A huge thank you goes to my family, I could not want for more from them. It is the ever present love and support from my parents that has shaped who I am today and I hope I make them proud. To Becca for being a great little sister and of course I must thank my brother, Ben, who has always paved the way forward for me. Your love and advice has helped me immensely. I would also like to thank my soon to be family, Nuala, Paul and Tess for all their support. I am lucky to be adding such great people to my extended family.

Finally, and most importantly, to Michaela. I was lucky enough to have you land on my doorstep in Manchester and we have now shared my PhD ups and downs. Your patience, support, belief, laughter and love helped me through. I am a better person for knowing you and I have no doubt that this is reflected in my work, thank you.

List of Figures

List of Tables

List of Equations

Research Abstract

In the post-genomic era a major paradigm shift occurred in biological research. The ability to catalogue all known genes and corresponding nucleotide sequences and to correlate this information with disease state, cellular function and phenotype, promised to change scientific approaches. As the wealth of genomic data increased, however, it became clear that far more information, in the form of the gene products (proteins), was required to put these new data into context. To achieve this, proteomics; the systematic analysis of the protein complement to the genome, became key. The protein sequence database information derived from gene sequencing heralded mass spectrometry-based proteomics emergence as an indispensable tool in biological studies. No other analytical tool offers the ability to determine the identity, quantity, modification status and cellular stability of hundreds to thousands of proteins in relatively short time periods. Continued technological and bioinformatics developments have increased the ability of mass spectrometry to catalogue proteomes and we are now at a point where close to comprehensive coverages are possible for simple organisms.

An ever increasingly aged western society has seen a steady increase in the prevalence of age-related diseases. Neurodegenerative diseases, are, like many other age-related diseases associated with protein misfolding. The cellular responses to protein misfolding and how they relate to disease clinical symptoms are not fully understood. It is, however, commonly accepted that protein misfolding is a key component in disease aetiology. Proteome alterations as a result of protein misfolding, therefore, represent an important avenue of research to better understand this debilitating category of diseases.

This work describes the use of label-free quantitative proteomics to expand upon currently known cellular responses to two misfolding diseases, Alzheimer's disease and α_1-antitrypsin deficiency. Initially a GeLC-MS^E workflow was developed and successfully utilised to identify proteins that were potentially implicated in Aβ42 toxicity and age-related vulnerability to Aβ42 in *D. melanogaster*. It is hoped that follow-up studies will provide more information on how these proteins are functionally implicated in the link between ageing and Alzheimer's disease.

Following this, a recently developed proteomic technology, that uses ion mobility to increase system peak capacity on the fly, HDMS^E, was employed to conduct time-course analysis of cellular responses to wild-type ageing and chronic Aβ toxicity in *D. melanogaster*. Using this more comprehensive approach, quantitative protein profiles across two lifespans were recorded. The data successfully identified large number of proteins that were significantly altered in abundance in relation to age and Aβ42 misfolding. The data represent a unique resource for follow-up reductionist and functional studies on the link between ageing and Alzheimer's disease.

Finally, a subcellular GeLCMS^E approach was successful at characterising ER enriched stress responses to the misfolding of two α_1-antitrypsin variants in relation to the wild-type, well folded and fully secreted variant. The present study expands on current pathways implicated in ER stress which have been largely characterised by reductionist studies. Interestingly co-fractionating bioenergetic components were quantified with large fold changes measured between variants. Follow-up live cell confocal imaging correlated these changes with mitochondrial elongation as a result of protein misfolding. This represents a novel finding not previously documented for α_1-antitrypsin deficiency.

Abbreviations

A1AT	α1-antitrypsin
AATD	α1-antitrypsin deficiency
Aβ	Amyloid Beta
AD	Alzheimer's Disease
ADP	Adenosine Diphosphate
APP	Amyloid Precursor Protein
B-H	Benjamini-Hochberg
BiP	Binding immunoglobulin protein
BLAST	Basic Local Alignment Search Too
CE	Collision Energy
CHO	Chinese Hamster Ovary
CID	Collision Induced Dissociation
CRM	Charged Residue Model
CV	Coefficient of Variance
Da	Dalton
DAVID	Database for Annotation, Visualization and Integrated Discovery
DC	Direct Current
DDA	Data Dependent Acquisition
DIA	Data Independent Acquisition
DTT	Dithiothreitol
E	Expectation
ECD	Electron Capture Dissociation
EDGE	Extraction of Differential Gene Expression
eIF	Eukaryotic Initiation Factor
EMRT	Exact Mass and Retention Time
EOR	Endoplasmic Reticulum Overload Response
ER	Endoplasmic Reticulum
ERAD	Endoplasmic Reticulum Associated Degradation

ESI	Electron Ionisation
ETD	Electron Transfer Dissociation
f	femto
FAB	Fast Atom Bombardment
FAD	Familial Alzheimer's Disease
FDR	False Discovery Rate
FTMS	Fourier Transform Mass Spectrometry
FWHM	Full Width Half Maximum
g	gram
GC	Gas Chromatography
GFP	[Glu^1]-fibrinopeptide B
GO	Gene Ontology
GRP-78	78 kDa Glucose-Regulated Protein
GS	Gene Switch
GSK	Glycogen Synthase Kinase
HIP	Hsc70-Interacting Protein
HPLC	High Performance Chromatography
hnRNPs	Heterogeneous Nuclear Ribonucleoproteins
Hsc70-3	Heat Shock 70-kDa Protein Cognate 3
Hsp	Heat Shock Protein
IAA	Iodoacetamide
iBAQ	Intensity Based Absolute Quantification
IDMS	Isotope Dilution Mass Spectrometry
IEM	Ion Evaporation Model
IM	Ion Mobility
iTRAQ	Amine-Reactive Isobaric Tagging Reagents
KEGG	Kyoto Encyclopedia of Genes and Genomes
L	Liter
LC	Liquid Chromatography
LOAD	Late-Onset Alzheimer's Disease

LOESS	Locally Weighted Linear Regression
LOPIT	Localisation of Organelle Proteins by Isotope Tagging
Min	Minute
M	Molar
m	Milli
MALDI	Matrix-Assisted Laser Desorption Ionisation
MDa	Mega Dalton
MCP	Microchannel Plate Detector
mol	Mole
m(t)RNA	Messenger (Transfer) Ribonucleic Acid
MOSE	Molecular Search Algorithm
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
m/z	Mass-to-Charge Ratio
n	nano
NEF	Negative Regulatory Factor
NHK	$Null_{Hong\ Kong}$
NMR	Nuclear Magnetic Resonance
Panther	Protein Annotation Through Evolutionary Relationship
PCA	Principal Components Analysis
PD	Parkinson's Disease
PDI	Protein Disulfide Isomerase
pI	Isoeclectric Point
PMF	Peptide Mass Fingerprinting
PLGS	ProteinLynx Global Server
PPI	Protein-Protein Interaction
ppm	Parts Per Million
PRM	Parallel Reaction Monitoring
PS	Presenilin
PTM	Post Translational Modification

QQQ	Triple Quadrupole Mass Spectrometer
Q-ToF	Quadrupole Time-of-Flight
RF	Radio Frequency
RP	Reverse Phase
RT	Retention Time
RU	Mifepristone
s/sec	second
SCX	Strong Cation Exchange
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SEM	Standard Error of the Mean
SILAC	Stable Isotope Labelling In Cell Culture
SOD	Superoxide Dismutase
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
SWATH	Sequential Isolation Windows
TBST	50 mM Tris, 150 mM NaCl, 0.05% Tween 20, pH 7.6
TCA	Tricarboxylic Acid Cycle
TMT	Tandem Mass Tags
ToF	Time-of-Flight
TP	Time Point
T-Wave	Travelling Wave
TWIMS	Traveling Wave Ion Mobility Mass Spectrometry
UGGT	(UDP)-glucose:glycoprotein glucosyltransferase
UPLC	Ultra Performance Liquid Chromatography
UPR	Unfolded Protein Response
V	Voltage
µ	Micro
3D	Three Dimensional
%	Percentage

1 Introduction

1.1 The Development of Mass Spectrometry as an Analytical Technique

Mass spectrometry is an analytical technique that is sensitive, specific and applicable to a multitude of workflows. For this reason, since its introduction by J.J. Thomson *et al* in 1912, mass spectrometry has been adopted as the preferred technique for the measurement of a wide range of analytes [1]. Proteins, metabolites, small molecule drugs and lipids to name a few are now routinely detected, characterised and quantified in academic and industrial settings by mass spectrometry.

In its simplest terms mass spectrometry is an analytical technique that allows the accurate measurement of an ionised molecule's mass-to-charge (*m/z*) ratio. Since its initial development in 1912 mass spectrometry has seen rapid advancement due to the versatility of the technique. J.J. Thomson's first experiments in 1897 identified that previously unquantified elements within cathode rays were deflected by a magnetic field to varying extents, which was later identified to be proportional to the particles *m/z*. Using the deflection properties of protons and electrons Thomson was able to ascertain that an electrons' mass is 1000 times less than a hydrogen atom. In one of his more famous experiments he was able to identify that neon consisted of more than one isotope, using the first purpose built mass spectrometer that he and Francis William Aston had built in the previous year. The identification of isotopes, which had been previously unknown, was based upon the observation of two distinct deflection patterns on photographic plates (Neon-20 & Neon-22).

Early applications of mass spectrometry focused on the analysis of small molecules with instruments that were either built by individual researchers or custom built by physicists interested in isotope determination and the study of ion formation. Early instrumentation employed magnetic-sector mass analysers. During World War II mass spectrometry was adopted for weapons-grade plutonium preparation. This led to the first ever commercialisation of such instrumentation. After WWII, time-of-flight (ToF) mass analysers, quadrupole mass analysers and gas chromatography (GC) for prior compound separation were developed. These developments made mass spectrometry applicable to areas such as organic chemistry and petroleum analysis which had been previously too challenging to analyse [2]–[5].

During the initial years of MS analysis ionisation of compounds for mass analysis was achieved either through electron ionisation (EI) or chemical ionisation (CI) [6], [7]. Although EI and CI are still used to good effect in small organic compound analysis, both ionisation methods have major limitations. The use of high energy electrons to produce molecular ions during EI often results in the complete fragmentation of the compound of interest. Subsequent identification of compounds after EI is often achieved by fragment ions giving a fragmentation pattern characteristic of the compound in question. CI is a lower energy technique that better preserves the original molecule intact and almost exclusively produces singly charged ions. The major drawback of the two ionisation methods is the inability to generate gas phase ions for larger biological structures such as large peptides and proteins that are both non-volatile and thermally labile.

The introduction of early desorption/ionization techniques were the first steps towards the analysis of large biomolecular complexes. Indeed fast atom bombardment (FAB), secondary ion mass spectrometry (SIMS) and plasma desorption were all significant steps towards the routine analysis of bimolecular complexes [8]–[10]. In particular FAB, which entails the bombardment of a glycerol-analyte solution with a high energy beam of non-reactive elements provided the first practical ionisation method for the study of biopolymers. The requirement of high concentrations of analytes due to inefficient ionisation and the inability to study large protein/DNA structures meant that mass spectrometry's applicability to biological sciences was still limited.

Electrospray ionisation (ESI) and matrix assisted desorption ionisation (MALDI) heralded mass spectrometry's establishment as a technique for the study of bimolecular molecules [11]–[13]. The impact was such that in 2002 John Fenn (electrospray ionisation) and Koichi Tanaka (soft laser desorption) received the Nobel prize in Chemistry for their work on the "development of methods for the identification and structure analysis of biological macromolecules". ESI in particular is still the main ionisation techniques chosen to help answer a variety of biological questions by mass spectrometry. These include protein identity, quantity, the presence of post translation modifications, subunit connectivity, topology and dynamics.

1.2 Electrospray Ionisation (ESI)

ESI, developed by Fenn *et al*, uses the generation of a fine liquid aerosol by electrostatic charging to generate gas phase ions, **Figure 1.1**. In electrospray, liquid is passed through a capillary held at a high voltage whilst a potential difference exists between the capillary and a counter electrode. The application of a high voltage (1-4 kV) causes charge to accumulate at the tip of the capillary. As charge accumulates the electrostatic Coulombic attraction to the counter electrode increases. As attractive forces increase beyond the point of the solvent surface tension, a spray forms in the shape of a pointed cone, termed the Taylor cone. This shape is formed due to droplets of cylindrical shape being able to hold more charge than when spherical. The droplets (µm in diameter) that constitute the Taylor cone evaporate whilst in flight, thereby reducing in size, resulting in a subsequent increase in charge per droplet volume. Similar to the action at the capillary, as charge density increases the applied electric field causes deformation of the droplets, eventual formation of a second Taylor cone and the ejection of smaller droplets. Each secondary droplet is around 1/10 the size of the ejecting droplet, making the ESI process efficient at producing a fine dispersion of liquid prior to large amounts of evaporation. It is from this population of highly charged small droplets that the majority of ions detected by mass spectrometry originate from.

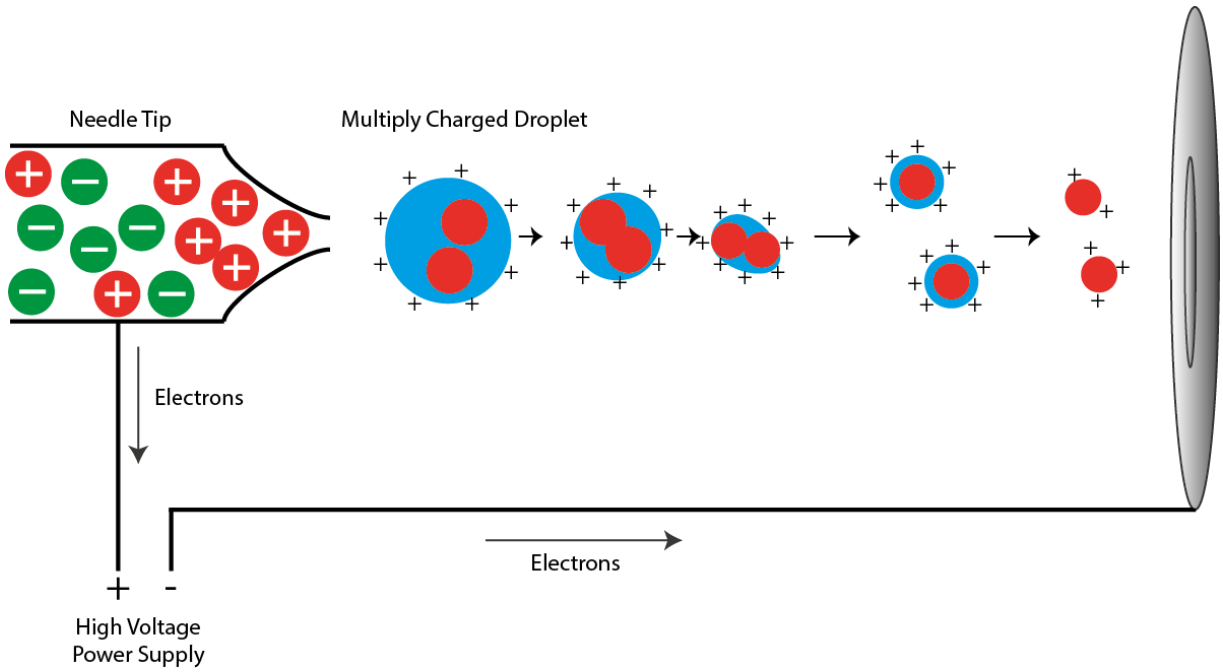


Figure 1.1 Diagram of the electrospray ionisation process

A number of possible ionisation mechanisms exist for ESI but the two predominant models are the ion evaporation model (IEM) and the charge residue model (CRM) [14], [15]. Ion evaporation is proposed for smaller analyte molecules and predicts that when droplet radii reach a certain limit, analyte molecules are desorbed from the surface, taking charge with them as they enter the gas phase. Alternatively, in the CRM, molecules, are ionised in a more passive fashion. In the CRM, which is thought to be more applicable to macromolecules, droplets undergo the evaporation-ejection process leading to the eventual situation where droplets contain an average of one or less analyte per droplet. Following eventual evaporation of all solvent, droplet surface charge is retained on the now gas phase analyte ion.

More recently a third mechanism has been proposed that involves both the IEM and CRM. This new mechanism states that IEM can occur before CRM and that the mode of ionisation is dependent on electric field strength at ion emission [16]. Ions with low requirements for electric field strength are emitted from the droplet via the IEM until depleted. Ions with higher requirements for field strength (larger analytes) are retained within the droplet and become gaseous ions via the CRM.

One of the beneficial characteristics of ESI is the production of multiply charged ions. If several ionisable sites are present then a number of these will be charged at any one time. During peptide and protein analysis this is observed as a charge state distribution. This is of particular advantage when analysing large molecules such as proteins, as it reduces the analyte *m/z* to a mass range that is optimal for data acquisition by commercial mass spectrometers.

1.3 Nanoelectrospray Ionisation (nESI)

ESI efficiency is largely governed by the size of the primary droplets. ESI flow rates of 1 ml – 10 µl min^{-1} mean that high temperature and pneumatic nebulization are required for adequate aerosol formation and efficient desolvation. Lower flow rates are employed during nESI, as developed by Wilm and Mann, (< 500 nl min^{-1}) and efficiently produces smaller droplets than ESI without the need for heat or sheath gas [17], [18]. nESI is performed with glass capillaries that have orifices with diameters in the range of 1-10 µm. For applications that require direct infusion the use a conductive material, e.g. gold, coated on the capillaries ensures adequate sample charging. For high-throughput

workflows, where MS analysis is coupled to chromatographic systems, charge can be applied at the nano-junction between column and capillary.

An obvious benefit of nESI, is the small sample amounts required for analysis. This has led to its widespread adoption in workflows where analysis is often sample limited, such as discovery proteomics, biomarker screening and metabolomics. Another benefit of nESI in such workflows is its increased tolerance for background ions such as salts. This is especially pertinent for quantitative proteomics where label-free quantification is sensitive to the matrix effects. The increased tolerance afforded by nESI originates from the production of smaller, but more highly charged droplets than those produced during ESI. Less droplet desolvation and fewer off-spring generations are therefore required during nESI prior to ion generation. As desolvation increases droplet concentration, reduced desolvation levels equates to reduced salt concentrations prior to ion formation.

1.4 Mass Analysers

Post ion production, ions are separated in relation to their *m/z* ratio where *m/z* analysis is based upon ion behaviour within an electric and/or magnetic field. Numerous mass analysers have been developed for the measurement of ion *m/z* ratios. The mass analyser chosen for use depends on both the application in question and the budgetary requirements of the laboratory. Broadly speaking there are two classes of mass analysers, analysers that perform quasi-simultaneous transmission/detection of ions, (time-of-flight (TOF), Orbitrap) and analysers that scan across an *m/z* region over a set period of time (quadrupoles, magnetic sectors).

1.4.1 Quadrupoles

Transmission quadrupoles or quadrupole mass filters are perhaps the most simple in design of common mass analysers currently in use. Quadrupoles, as the name suggests, consist of four hyperbolic rods arranged symmetrically, **Figure 1.2.** Opposing rods are connected electrically and a radio frequency (RF) voltage is applied between each pair of rods. A direct current is superimposed on the RF voltage. At any given RF voltage only ions of a particular *m/z* ratio range will have a stable

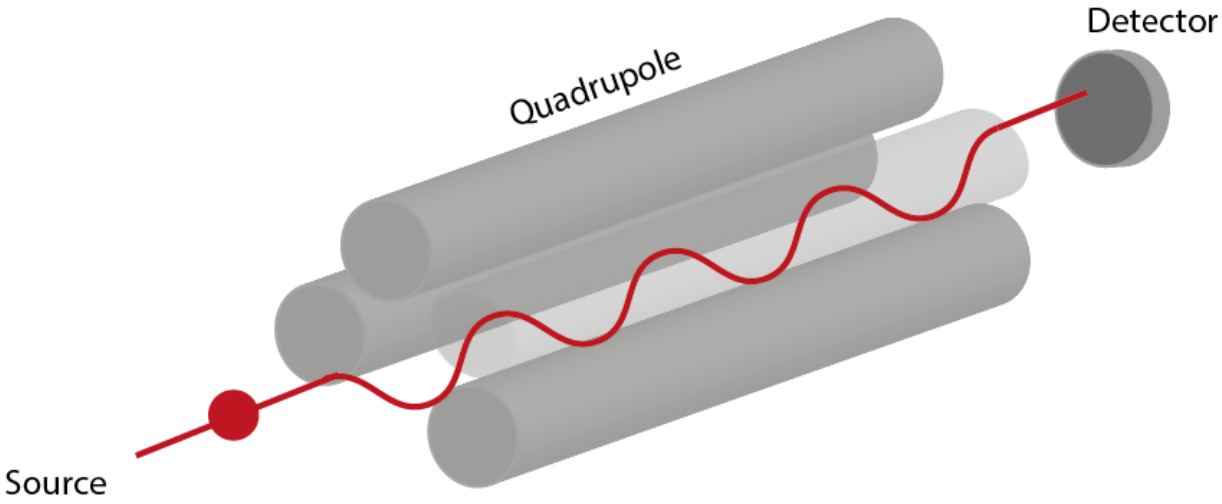


Figure 1.2 Representation of stable ion path through a quadrupole mass analyser

trajectory and successfully reach the detector. Unstable ions will collide with the rods and not be detected. Based on a required transmission window (*m/z* range), the required RF and DC voltage components can be calculated using the reduced Mathieu equation, **Equation 1.1-1.3**, where q_x and a_x are used to calculate the amplitudes of both components. Here z is the ion charge, e is the total charge of an electron, r_0 is half the distance between opposing rods, m is the ion mass and ω is the angular frequency of an ion. The stable path of a particular ion is defined by the magnitude of the RF voltage V and by the ratio U/V where U is the direct current magnitude.

Equation 1.1

$$a_x = -a_y = \frac{4zeU}{m_i r_0^2 \omega_2}$$

Equation 1.2

$$q_x = -q_y = \frac{2zeV}{m^2 r^2 \omega^2}$$

Equation 1.3

$$\tau = \frac{\omega t}{2}$$

In order to obtain a full mass spectrum, quadrupoles can be set to scan across a chosen *m/z* range, **Figure 1.3**. The number of steps used during the scan is proportional to spectrum mass resolution

and inversely proportional to duty cycle. For this reason, although the maximum resolving power of a typical quadrupole is ~ 3,000, they are most often operated in unit mass resolution, representing the ideal trade-off between required resolution and duty cycle. Alternatively quadrupoles can be used in RF mode only, where they act as high-mass filters to transfer ions between regions of mass spectrometers.

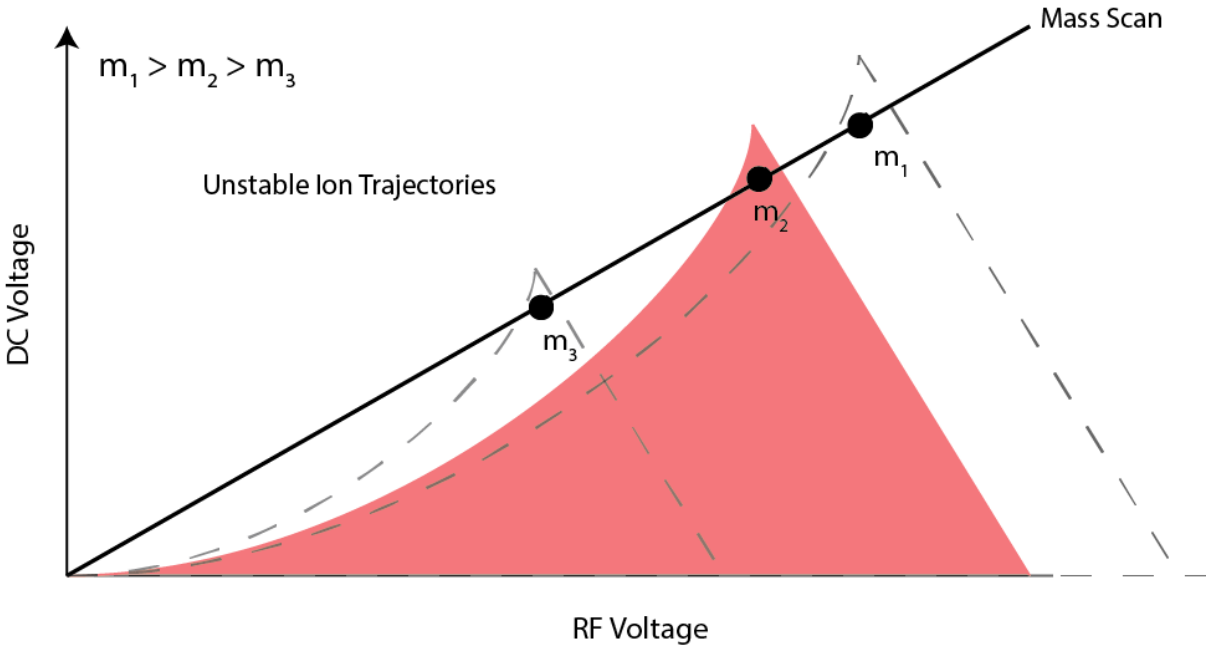


Figure 1.3 Ion stability graph as a function of DC and RF voltage. Triangles indicate stable regions for ions of different masses. Solid line indicates the quadrupole mass scan that would successively transmit the three different ions

1.4.2 Time of Flight (ToF)

The first published design of a linear ToF mass spectrometer was presented in 1952 by William E. Stephens [19]. All ToF mass analysers operate on the principle of measuring the time required for an ion to travel from an ion source to a detector. For this reason, unlike quadrupole mass analysers that operate with a continuous flow of ions, TOF mass analysers require pulses of ions to measure *m/z.*

During ToF mass acquisitions, packets of ions are accelerated by a potential difference V, within an acceleration region. Ion packets are either generated by an intermittent ionisation process, such as MALDI, or if coupled to a continuous ionisation source by the transient application of potential energies. The application of a constant potential results in all ions acquiring the same amount of

kinetic energy. Ions leaving the acceleration region enter a field free region, commonly referred to as the flight tube, which is maintained at a high vacuum (E^{-7} Torr). By measuring the time taken for each ion to reach the detector, which is positioned at the end of the field free region, the *m/z* of an ion can be calculated using **Equation 1.4-1.8**

As mentioned all ions receive the same kinetic energy in the acceleration region which is given by:

Equation 1.4

$$qV = zeV = E = \frac{1}{2mv^2}$$

Where m = ion mass, q = total ion charge, V= push voltage, z= ion charge, e= electron charge, E= total energy and v= ion velocity. Further to this the time taken t for an ion to traverse the field free region is given by:

Equation 1.5

$$t = \frac{s}{v}$$

Where s = distance and v = ion velocity. Upon substitution of v with **Equation 1.4** gives:

Equation 1.6

$$t = \frac{s}{\sqrt{\frac{2ezV}{m_i}}}$$

Finally rearrangement of Eq. reveals the relationship between time taken for an ion to reach the detector and its *m/z*

Equation 1.7

$$t = \frac{s}{\sqrt{2eV}}\sqrt{\frac{m_i}{z}}$$

With the knowledge that $\frac{s}{\sqrt{2eV}}$ is a constant, it is clear that t and *m/z* are directly proportional. Calibration of a ToF mass analyser with chemicals that yield ions of known *m/z* ratios can be used to calculate the constant within **Equation 1.7.**

Linear ToF analysers offer a high level of sensitivity as transmission efficiency is close to 90 % with low levels of losses caused by collisions with residual gases. Due to the quasi-simultaneous detection of ions, ToF analyser also offer much faster acquisition speeds than quadrupole analysers. A spectrum can be obtained across a large *m/z* region in a µsec time period. Another benefit of ToF analysers is the high mass detection limits afforded, which, given an infinitely long flight tube is theoretically limitless. Practically ToF analysers offer the highest mass range of all commercially available mass analysers. Indeed in 2013 a ToF analyser was used by Snijder and co-workers during the native MS analysis of a 18 MDa virus, highlighting the ability of ToF instrumentation to detect very large ions [20]. The main limitation of linear ToF analysers is the low mass resolution that is obtainable. Resolution is limited due to a number of factors. Theoretically all ions prior to acceleration have zero kinetic energy and are focused into the same region with little spread. This, however, is not true in practice with size of the ion volume, length of ion formation pulse (for pulsed ionisation methods) and initial kinetic energy variations all contributing to reduced mass resolution, **Figure 1.4**. The most obvious method for increasing the resolving power of a ToF analyser is to increase the length of the flight tube, but this however can become impractical for use in laboratories. Expanding upon **Equation 1.7** shows the direct relationship between flight tube length and mass resolution.

Equation 1.8

$$\frac{m}{z} = \left(\frac{2eU}{S^2}\right) t_2$$

and integrating

Equation 1.9

$$\frac{1}{z}\, dm = \left(\frac{2eU}{S^2}\right) 2t\, dt$$

Gives

Equation 1.10

$$\frac{m}{dm} = \frac{t}{2dt}$$

Equation 1.11

Finally

$$R = \frac{m}{\Delta m} = \frac{t}{2\Delta t} \sim \frac{S}{2\Delta x}$$

Where m is mass, t is flight time, Δm and Δt are the peak width at ½ levels on the mass and time scales respectively. Δx denotes the ion packet thickness as it approaches the detector.

The use of delayed pulse extraction, where a time lag between ionisation and extraction is introduced, overcomes differences in initial ion kinetic energies [22]. Ions in the source are allowed to expand into a field-free region for a period up to several microseconds before an extraction pulse is applied. The result of this delayed pulse is that ions with more initial kinetic energy are further from the source of acceleration when the pulse is applied and therefore receive less energy than ions which were closer to the source.

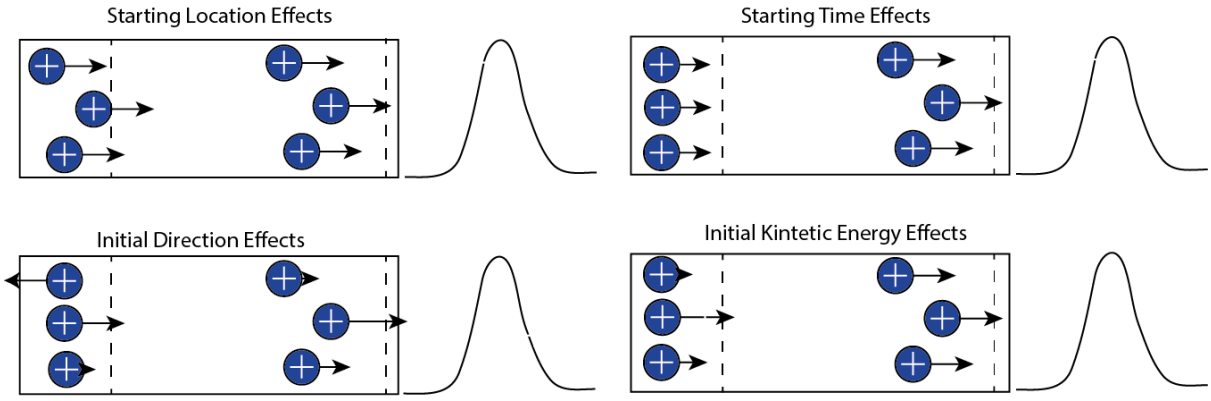


Figure 1.4 Time, space and kinetic energy distributions of ions during time of flight that cause of reduction in data mass resolution. Adapted from [21]

Even with delayed pulse extraction the kinetic energy dispersion of ions limits the resolution of linear ToF analysers. The use of an electrostatic reflector, more commonly known as a reflectron is an effective way to reduce the effects of kinetic energy dispersion, thereby improving mass resolution [23]. In its most simple form a reflectron consists of a retarding electric field located opposite the ion source and at the end of the field-free region, **Figure 1.5**. The electric field is composed of a series of electrodes (or rings) at an increasing potential. Ions entering the field are slowed by the field until they reach zero kinetic energy at which point they are accelerated out of the field in the opposite direction they entered. Ions with more kinetic energy will travel further through the reflectron than lower energy

ions. This path length differential corrects for time-of-flight differences due to kinetic energy dispersion and dramatically improves mass resolution. Using this approach current reflector ToF analysers can achieve a mass resolution > 25,000 full width half mass (FWHM). If more than one

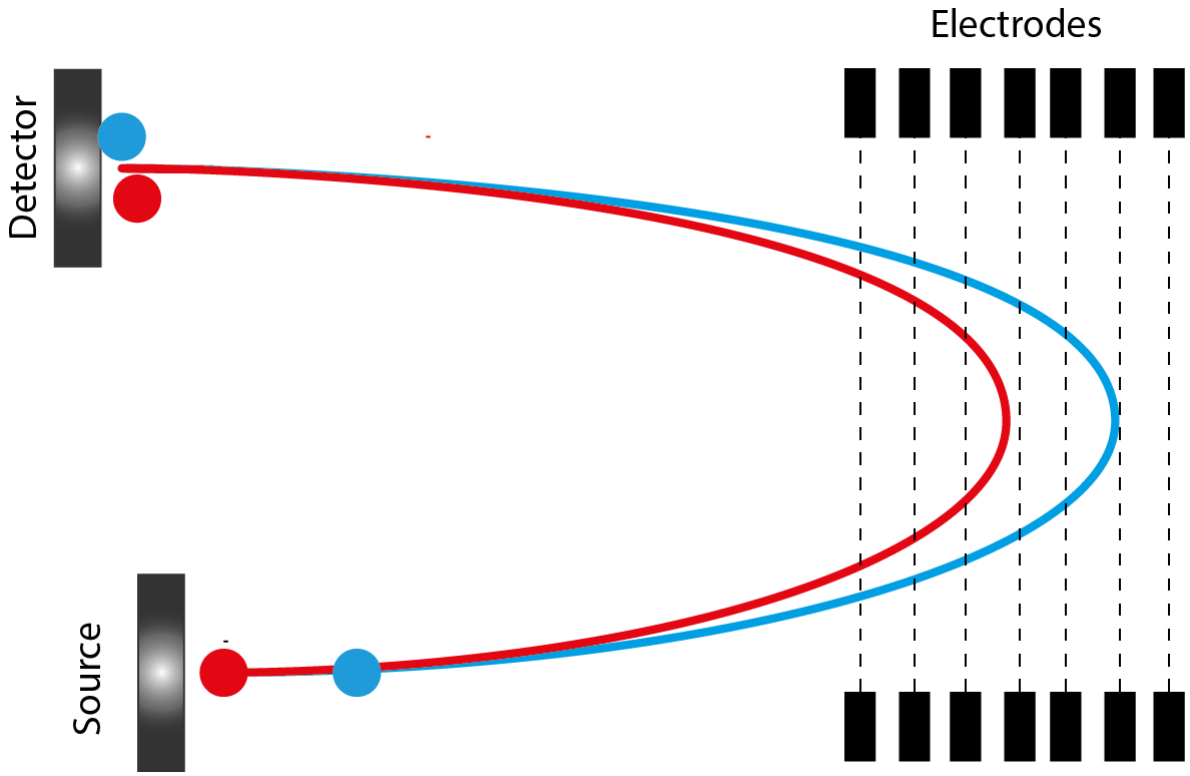


Figure 1.5 Schematic of a reflectron time of flight mass analyser. Shown are two ions of the same *m/z* but with different kinetic energies. Here the blue ion has more kinetic energy than the red ion.

reflectron is employed higher still mass resolution can be achieved but at the cost of reduced ion transmission.

1.5 Detectors

As ions leave the mass analyser, they must be detected, most commonly as an electrical signal. There are various methods of detection including photographic plates, photon multipliers and array detectors but the preferred method is electron multiplication.

An electron multiplier uses the property that when energetic particles collide with a metal or semiconductor, secondary particles (positive ions, negatives ions, electrons and neutrals) are emitted.

The type of particle emitted from the electron multiplier conversion dynode is dependent on the particle that originally collided. In an electron multiplier this property is enhanced by holding the conversion dynode at a high potential, opposite the charge of the particle to be detected. This accelerates the ions leaving the mass analyser. High velocity impacting ions cause higher numbers of subsequent emissions thereby increases the sensitivity of detection. By placing an electrode, or primary dynode at a more positive potential opposite the conversion dynode, emitted electrons are accelerated towards the dynode. Upon impact, further emissions occur, thereby amplifying the number of emitted electrons. This process is repeated a number of successive times, with dynodes arranged in a series, until the current is measured. This process can result in an amplification of signal up to 7 orders of magnitude.

In modern mass spectrometers, the detector most commonly used is a microchannel plate detector (MCP). A MCP consists of an array of parallelised electron multiplier channels. Electron multiplication is achieved by coating each channel with a semiconducting material and the curvature of each channel is designed to prevent positive ions accelerating toward the input side. Amplification achieved by a MCP detector is around 10^3-10^4 and for this reason two are often used in conjugation. A benefit of MCP detectors is that due to the large number of channels, many different ions can be detected simultaneously, where a single ion that reaches the detector will only affect a small number of channels. The large detection area of the plate also removes the requirement for additional ion focusing prior to detection. A major limitation of MCP detectors is that a finite number of ions can be detected before detection saturation is reached and further ion collisions are not detected. Saturation is a result of the intrinsic time period taken for each channel to recover from an ion detection event before further ions can be detected. This dynamic range limitation can prevent low abundance ions being detected if they are co-ionised with a much higher abundance ions, a common occurrence during the analysis of complex biological samples.

Previously time-to-digital converters (TDC) were used to digitise a ToF signal recorded by MCP detection. A TDC acquisition requires a trigger (or discriminator), where once a signal threshold is met, data recognition will begin. As the threshold is met, the signal in question is registered by a counter which records flight time. A brief dead time then follows before the counter and discriminator are ready to record the next ion that arrives with above the required intensity. A TDC approach limits

dynamic range due to its ability to only record the initial incoming ion signal. TDCs mark the arrival of an ion but do not record how many ions were the result of this mark. When multiple ions of a given *m/z* are present during a ToF push or transient TDC detection underrepresents the true number of ions that are present. Improvements in ion optics and ion sources means that current mass spectrometry sensitivity is high enough that TDC systems do not offer a large enough dynamic range for quantification. The use of analogue-to-digital converters (ADC) is now commonplace the systems offer a much higher linear signal dynamic range. An ADC does not determine the exact arrival time of ions, it only records detector output at a fixed interval. During each cycle of detection signal is converted to digital data with data from successive cycles being combined. As an ADC is an analog device it is able to combine signal from several near simultaneous ion arrivals thereby providing data representative of ion counts per ToF push.

1.6 Tandem Mass Spectrometry (MS/MS)

Tandem mass spectrometry is the coupling of two mass analysis stages. Between the first and second mass analysis there is either a dissociation process or chemical reaction that causes a change in the mass or charge of an ion, **Figure 1.6A**. The data obtained can be used to determine more chemical or structural information on the ion in question than is possible from MS analysis alone. Tandem mass spectrometry can be achieved in two ways, either in space or in time, with the former being the most common approach. MS/MS in space involves the coupling of two distinct mass analysers together, whilst MS/MS in time is achieved by sequentially performing mass analysis steps in an ion storage device. In order to perform tandem in time, an analyser that is capable of trapping ions is therefore required. Proteomics makes use of a number of different tandem mass spectrometry configurations to study proteins and peptides including the Q-TOF (a hybrid instrument where a quadrupole is coupled to a ToF analyser), the triple quadrupole, QQQ, (three quadrupoles placed in sequence) and the ion trap. A widely used ion trap-based instrument is the Orbitrap, a modified ion trap where an ion's radial trajectory around a center electrode is directly related to its *m/z* [24]. Deconvolution of data is performed by Fourier transform. An Orbitrap instrument is an unusual mass analyser as although all mass analysis steps can be performed in the Orbitrap itself, ion dissociation is performed in a second ion trap placed prior to the Orbitrap mass analyser.

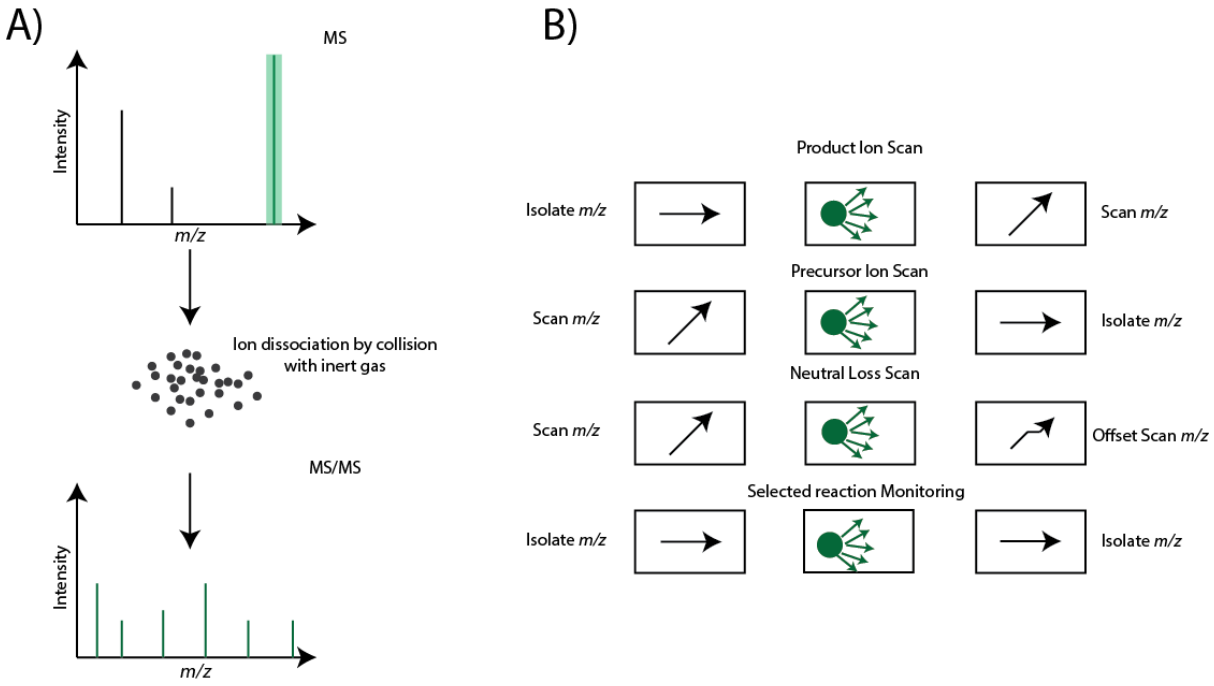


Figure 1.6 Tandem mass spectrometry. **A)** Schematic of precursor ion isolation and dissociation **B)** Four common MS/MS modes of acquisition

Depending on instrument configuration, there are four main MS/MS experiments that can be performed with each experiment providing distinct information on the given ion, **Figure 1.6B**. During a product ion scan, a specific precursor ion is isolated for dissociation. Following this, an *m/z* scan is performed to collect information on the product ions produced. This experiment is commonly performed in proteomic studies used to characterise tryptic peptides present in a sample. A precursor ion scan consists of choosing a product ion of interest and determining the precursor ion/s whose dissociation results in that ion's formation. This is achieved by using the first mass analysers to scan across the precursor *m/z* range whilst the second mass analyser is focused on the product ion *m/z*. The third experiment that can be performed is the neutral loss scan. The aim of a neutral loss scan is to identify all precursor ion dissociations that result in a characteristic neutral loss. To do this, both mass analysers are operated in scanning mode, the second analyser, however, is offset by the neutral loss being searched for. By operating in this manner, detection of an ion occurs if the precursor and product ions have a mass difference equal to that of the neutral loss. A fourth, but not final mode of operation possible is selected reaction monitoring (SRM), an experiment most often used for ion quantification due to the high specificity and sensitivity of the method [25]. In a SRM

experiment the first analyser isolates a precursor ion. Following dissociation of the ion, the second mass analyser isolates a single product ion. Sensitivity is increased in this mode as both analysers are focused on a particular *m/z* range; increased times spent on the chosen detection range increases sensitivity.

1.7 Collision Induced Dissociation (CID)

As mentioned, a tandem mass spectrometry experiment involves either a dissociation process or chemical reaction between the two mass analysis steps. The most widely used of the dissociation techniques is collision induced dissociation (CID). CID is a fragmentation process whereby ions are accelerated through a chamber that contains an inert gas (He, N_2 or Ar) maintained at a higher pressure than the surrounding vacuum. The gaseous ions collide with inert gas molecules, in a very fast process (10^{-14} – 10^{-16} s) where a fraction of the ion translational energy is converted into internal energy. The faster the ion is accelerated, the more energetic the collisions that occur are. The ion, now in an excited state undergoes unimolecular decomposition. Ion disassociation is slow in comparison to the collisions and allows the randomisation of ion internal energy. Subsequent fragmentation only occurs at bonds that are lower in energy than the redistributed internal energy. This 'ergodic' fragmentation pathway is particularly beneficial for the analysis of peptides as it yields predictable fragment ions. A product ion nomenclature, initially proposed by Roepstorff and Fohlman[26] but later modified by Biemann,[27] has been adopted for peptide analysis because of the uniform fragmentation pathways that are observed, **Figure 1.7**.

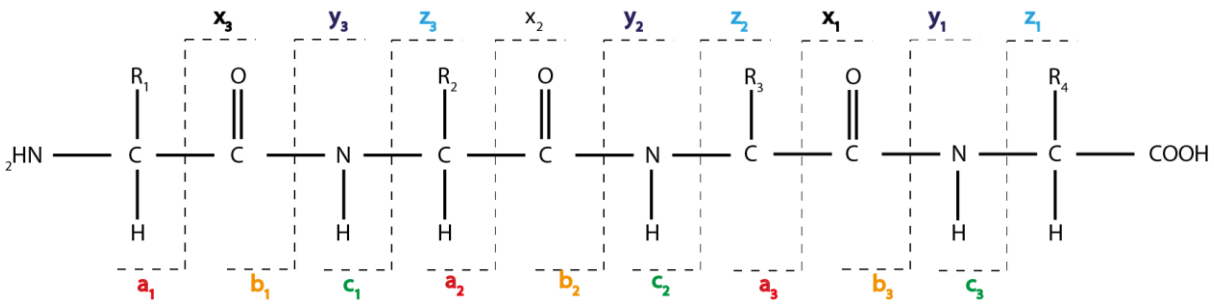


Figure 1.7 Fragmentation ion nomenclature for peptides

In CID molecules fragment at the weakest bonds of the ion being analysed. For peptides this is the amide bond, yielding what are referred to as b-and y-ions. b-product ions are formed when, during fragmentation, charge is retained on the N-terminus fragment, whilst y-ions are formed by retention of charge on the C-terminus fragment. Both sets of ions are numbered from the most terminal fragment ion. The site of fragmentation for a protonated peptide is best described by the mobile proton theory which states that immediately following ionisation positive charge will be retained on the most basic amino acid residues. Following ion activation, however, proton[s] can be transferred from the less-basic of the initially occupied sites and cause subsequent charge-site initiated fragmentation. It is, therefore, the position of the mobile proton that influences the site of peptide bond dissociation.

A major caveat of CID is that due to the most labile bonds being fragmented, its applicability to the analysis of post translation modifications (PTMs) is limited. PTMs, such as phosphorylation, have labile bonds and are therefore lost upon CID. Subsequently it is possible to identify phosphorylated peptides using CID, however if two or more possible sites of phosphorylation are present it is not possible to locate the exact site of modification. For the analysis of PTMs, electron capture dissociation (ECD) and electron transfer dissociation (ETD) are better suited.

1.8 Electron Capture Dissociation (ECD) & Electron Transfer Dissociation (ETD)

Electron capture dissociation (ECD) and electron transfer dissociation (ETD) are analogous dissociation techniques that transfer electrons to positive ions to create radical cations following charge state reduction.[28], [29] The radical cations formed then undergo fragmentation. The difference between the two fragmentation techniques is the method of electron transfer, with ECD using low energy electrons whilst ETD utilises radical anions such as anthracene or fluoranthene.

During dissociation by ECD, electron capture by a multiply charged ion occurs with charge state reduction to form an odd electron cation. This increases the ion's internal energy leading to fragmentation. The process of activation is very short ($< 10^{-14}$s) and is thought to occur faster than bond vibration, meaning fragmentation is directed by radical ion chemistry. ETD fragmentation follows a similar pathway, with the main difference being the method of electron transfer. As such both ECD and ETD are non-ergodic processes.

The use of a strong electric field during ECD makes it incompatible with the majority of mass analyser and must be combined with a Fourier transform ion cyclotron resonance mass analyser (FTICR). For the ECD process to be efficient, precursor ions must be maintained in a dense population of near-thermal electrons. The RF fields used in ion traps and quadrupole time –of-flight instruments do not maintain the thermal energy for periods long enough to induce sufficient fragmention. ETD was developed to circumvent this problem. Reagents such as fluoranthene can be delivered to the source whilst a high voltage generates the required anions.

ECD and ETD fragmentation offer complementary information to CID. Dissociation occurs about the amine bond, producing primarily c and z-type ions, **Figure 1.7**. It has previously been shown that employing both CID and ETD in a proteomics workflow can provide better peptide sequence coverage than the two fragmentation methods provide separately [30], [31]. Where ECD and ETD offer the most benefits however, are in the analysis of post translation modifications and larger peptides/proteins, with, ETD in particular being heavily adopted in these areas [30]–[34]. With the ETD fragmentation being driven by radical ion chemistry, labile bonds such as PTMs, which are lost during CID, are maintained, allowing the determination of site specific modifications. For larger peptides and proteins, ETD has been demonstrated to provide better sequence coverage than CID [35]. ETD's efficiency is seen to increase with increasing ion charge, a correlation that is explained by Coulombic effects. Larger peptides and proteins contain multiple basic sites for protonation and tend to be highly charged post ESI which is beneficial for fragmentation efficiency. Indeed the fragmentation of large biomolecules by ETD has been demonstrated to provide better sequence coverage than CID. ETD has yet to be universally adopted for discovery-based proteomic studies mainly due to the lower efficiency fragmentation provided by ETD in comparison to CID when analysing short peptides (~10 residues) typically produced by a tryptic digestion.

1.9 The Emergence of Mass Spectrometry-Based Proteomics

After the development of rapid DNA sequencing methods in the 1970's, advances in whole genome sequencing technology have facilitated the determination of the entire genetic makeup of complex eukaryotic organisms including humans [36]. Proteomics is now benefiting from archived and annotated sequence information that allows results to be queried and correlated to distinct biological

pathways. The term proteome was coined in 1975 and refers to the protein complement to the genome [37]. Proteomics is, therefore, the experimental analysis of the proteome; the complete protein content of a defined set of cells. The complexity of the proteome and its dynamic nature limits the ability of current mass spectrometric instrumentation to conduct proteomic analyses without experimental bias. Proteomic analyses, therefore, tend to fall under a number of categories depending on the desired target of the experimental design. Protein translation, turnover and post translational modifications are all areas proteomics is able to characterise to better understand how complex biological processes function through a systems level approach.

Mass spectrometry is now the technique of choice for proteomic investigations due to its sensitivity (low fmol-amole), versatility and its high throughput nature. Early work in proteomics focused on the identification and quantification of a select number of proteins, however with rapid advancement in both mass spectrometry instrumentation and chromatographic systems this number has increased by orders of magnitude. Whereas immunoassay techniques such a western blots and enzyme-linked immune sorbent assays are able to quantify small numbers of target proteins with reasonable sensitivity within a single experiment, mass spectrometry is now able to identify and quantify up to 4000 proteins in a single hour analysis **Fig 1. 8** [38]. The antibody free nature of mass spectrometry-based analysis provides a major benefit in relation to the alternate immunoassay-based techniques. This allows a far more unbiased approach to cell biology studies to be carried out. Indeed a study reported by Edwards and colleagues compiled a selection of research papers before and after the complete human genome was sequenced [39]. The findings of the study were two-fold; firstly the majority of researchers are focused on a very small percentage of proteins, which is directly related to the availability of effective antibodies and secondly that increased accessibility to human genome data and the proteomic techniques has done little to change this trend. An additional benefit to mass spectrometry is that, if performed correctly, added specificity can be provided. The large number of variables that a mass spectrometry-based proteomic experiment records, including peptide(s) retention time, peptide(s) mass and peptide fragment(s) mass in order to identify a protein offers a specificity not possible by antibody based techniques [40].

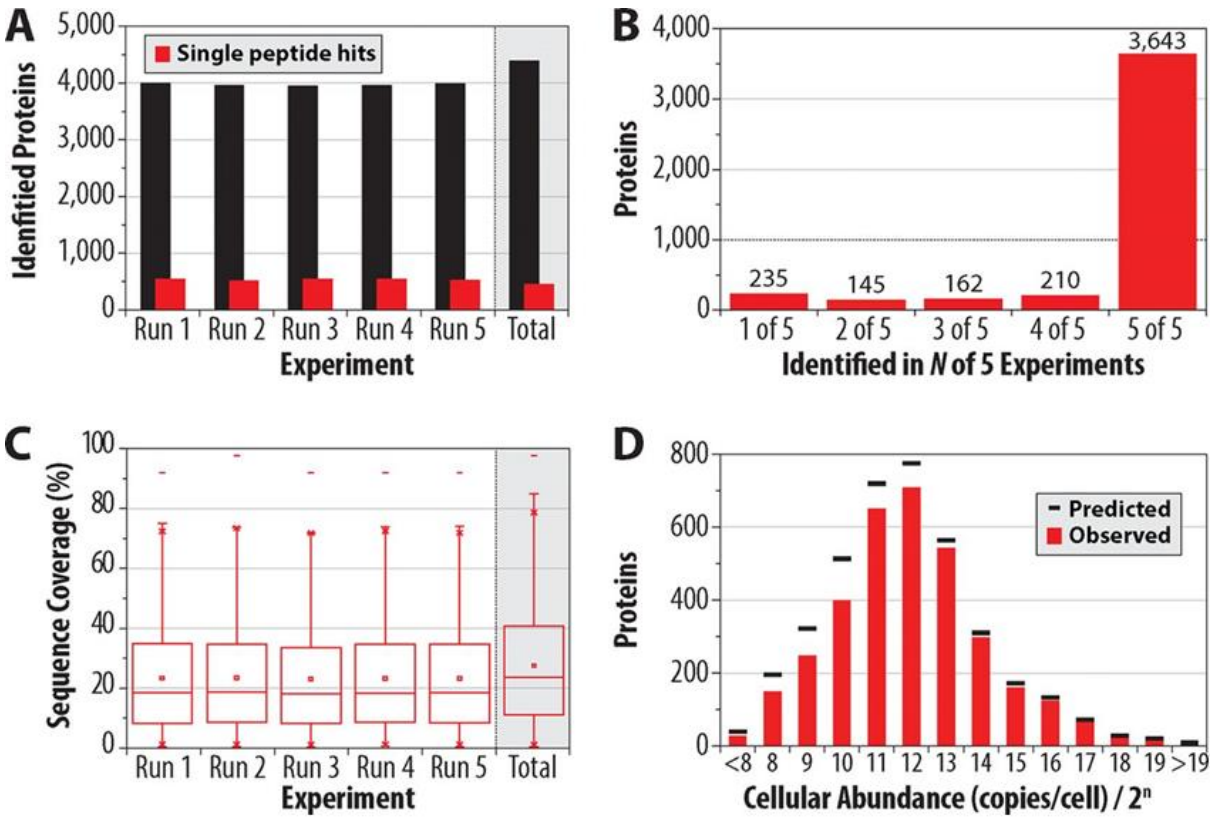


Figure 1.8 The one hour yeast proteome. Performance metrics from the analysis of a yeast tryptic digest using a 70 min LC gradient analysed on an Orbitrap Fusion. **A)** Protein identifications **B)** Protein identification reproducibility **C)** Box plot of protein sequence coverage **D)** Dynamic range of proteins identified. This research was originally published in MCP. A. S. Hebert *et al,* The one hour yeast proteome, *MCP,* 2013, 13(1), 339-47. © the American Society for Biochemistry and Molecular Biology[38]

1.10 Proteomic Workflows

Mass spectrometry-based proteomics is a wide-reaching subject matter and the aims of the experiment in question guide the workflow chosen. Despite this the majority of proteomic experiments follow a very similar protocol with an ever increasing number of variations on this model possible, **Figure 1.9**. The generic workflow consists of protein extraction from the system being characterised. Following protein extraction into a suitable buffer, protein level separation can be used to decrease sample complexity. Protein samples are then proteolytically cleaved into peptides. Enzymes such as trypsin are preferred due to the specific nature of cleavage and the generation of peptides that upon ionisation have an *m/z* in the range suited to mass spectrometric analysis. MS instrumentation is

commonly coupled to reverse phase- high pressure liquid chromatography (RP-HPLC) systems for peptide level separation prior to analysis. If further separation is required, orthogonal peptide separation techniques (e.g. strong cation exchange) can be employed prior to RP separation. Mass spectrometry analysis can be acquired in a non-targeted (discovery proteomics) or targeted manner (quantitative validation) and the choice depends on both experimental aims and instrumentation available. Data acquired are then searched against a protein database. At this stage there are a number of different software programs available for this task. Database searching for peptide identification is possible due to the predictable nature of peptide fragmentation and the specificity of enzymatic cleavage. Protein identification is then inferred from the presence of characteristic peptides predicted from *in-silco* digestion.

1.11 The Need for Prior Separation

The proteome is a far more dynamic and complex entity than the genome. Whilst the human genome contains ~20,000 protein coding genes, splice variants, single-nucleotide polymorphisms (SNPs) and PTMs means the number of protein variants (isoforms and proteoforms) translated from the genome is far higher. This complexity has challenged proteomics and driven mass spectrometry instrumentation developments. A typical sample obtained from cell lysis for proteomic analysis contains thousands of distinct proteins. Enzymatic digestion before analysis increases complexity by at least an order of magnitude. It is for this reason that MS instrumentation is coupled with chromatographic equipment, primarily reverse phase liquid chromatography. Separation of peptides based on their physicochemical properties reduces sample complexity prior to MS analysis but does not ensure the detection of all peptides present in a sample. In a research paper presented by Michalski and colleagues it was shown that during the analysis of a HeLa cell lysate, an estimated 100,000 peptide like species were observed by LC-MS [41]. Peptides targeted for MS/MS were however only 16 % of the total number of peptide-like ions, therefore dramatically limiting the

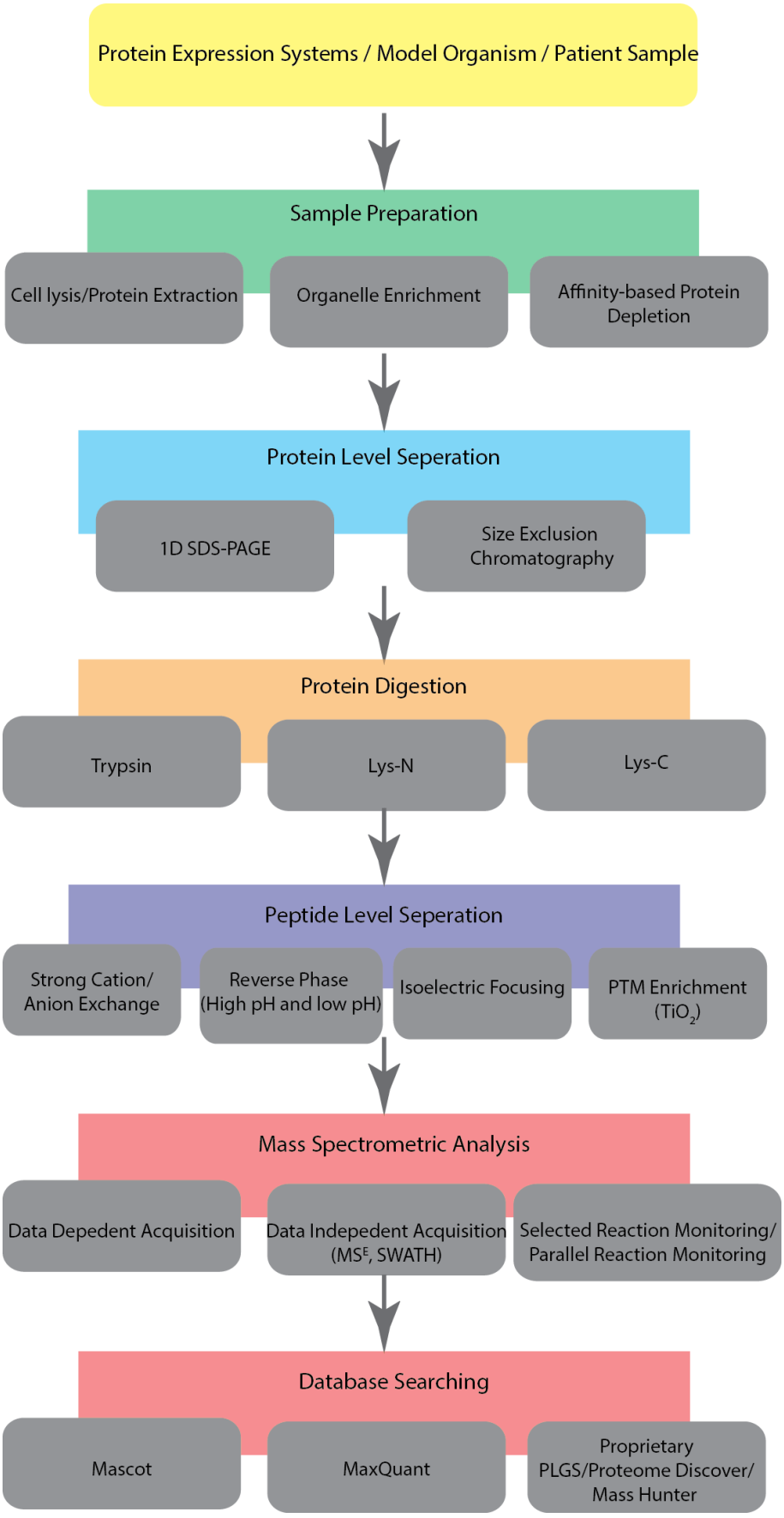


Figure 1.9 Generalised proteomics workflow. At each stage common techniques/software programs are indicated

number of peptides successfully sequenced. Instrumentation scan speed and sensitivity has since improved, increasing the number of peptides accessible by MS, however proteomics is still limited by sample dynamic range. Modern MS instrumentation has a detection dynamic range of up to four orders of magnitude. This can be increased by a further order if targeted analysis is undertaken. Protein abundances span at least seven orders of magnitude increasing to 10 in the case of human plasma. It is this discrepancy that limits mass spectrometry proteome coverage to the identification more abundant proteins, **Figure 1.10**. In a recent review of the challenge that the proteome dynamic range presents discovery proteomics, Zubarev examines the proteome abundance distribution in respect to mass spectrometer dynamic range. He subsequently concludes that efforts should be focused on developing methods as an alternative to extensive fractionation to enable rapid deep proteomics [42].

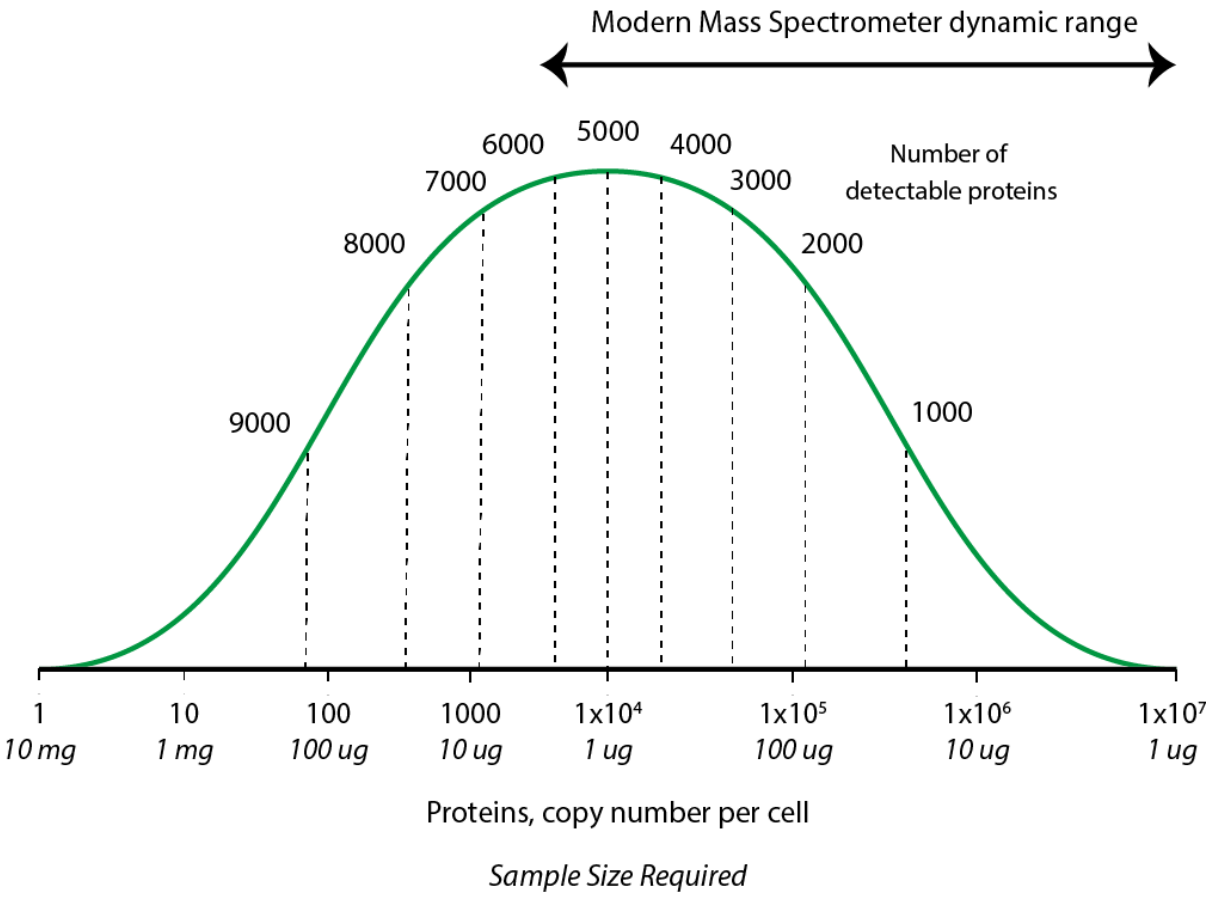


Figure 1.10 Density plot of protein numbers in reference abundance (copy number per cell) for a typical eukaryotic cell lines or tissue sample. Image adapted from R. A. Zubarev, 2013. [42]

1.12 Peptide Level Chromatographic Separation

1.12.1 High Pressure Liquid Chromatography (HPLC)

High pressure liquid chromatography is the most commonly employed on-line fractionation method coupled to MS. Reverse phase separation is used to separate peptides based on their hydrophobicity and is well suited to MS as the solvents are directly compatible with ESI. For proteomics research, separations are performed on nano columns with an internal diameter (ID) ≤ 100 µm, designed to provide separations for sample amounts of < 1 µg. Due to sample complexity in proteomics the peak capacities (P_c) provided by a typical HPLC column are not sufficient to adequately separate all peptides[43]. Solutions to increase column resolution typically include increasing the length of the column and/or decreasing the particle size[44], [45]. Both solutions increase column stationary phase surface area but at the expense of increased back pressure. Typically HPLC systems can operate up to ~ 6000 psi which limits the degree to which one can alter the two parameters and therefore limits maximum peak capacity.

1.12.2 Ultra High Pressure/ Ultra Performance Liquid Chromatography (UHP/UPLC)

Ultra performance liquid chromatography (UPLC) was made commercially available by Waters Ltd in 2004. The first ACQUITY UPLC systems were initially capable of operating at back pressures of up to 10,000 psi which opened up the possibility of utilising significantly decreased particle sizes. With the introduction of the M-class ACQUITY range the back pressure limit has been increased further to 15,000 psi. It can be seen from the van Deemter plot, based upon the empirical formula, **Equation 1.12,** that describes the relationship between flow rate and column efficiency, **Figure 1.11**, that there are additional benefits of sub 2 µm particles beyond increased peak capacity[46]. Here A is the Eddy-diffusion parameter, B is the diffusion coefficient, C is the resistance to mass transfer coefficient and u is the linear velocity of the solvent. At larger particle sizes it can be seen that there is a limited optimal mobile phase flow rate to achieve maximal peak resolution. Beyond this flow rate resolution rapidly deteriorates and this characteristic is more extreme for the larger particles. Sub 2 µm particle-based columns do not have this problem and significantly higher flow rates can be tolerated without a large

reduction in peak resolution. This allows the development of high throughput methods where good peak separation can be achieved with short run times. This attribute is of particular benefit in targeted validation of discovery proteomic findings where a large number of analytical runs are typically required.

Equation 1.12

$$Resolving\ power = A + \frac{B}{u} + (C_s + C_m)\ x\ u$$

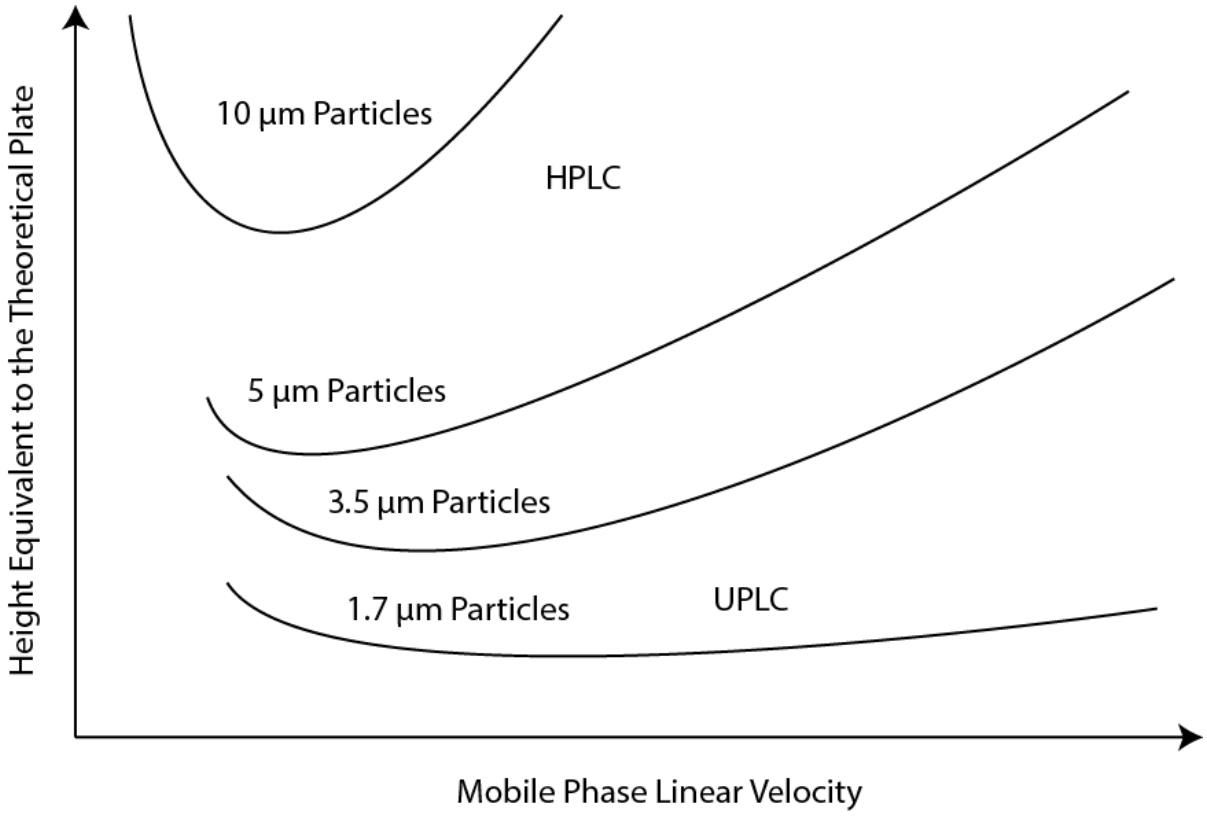


Figure 1.11 Van Deemter curves for four common particle sizes. Adapted from L. Novakova, 2006

1.12.3 Multidimensional liquid chromatography (MD LC)

Multidimensional liquid chromatography is the combination of two or more forms of LC to offer a higher degree of chromatographic peak capacity in order to increase peptide identification rates[47]. The LC methods coupled together have, to some degree, orthogonal peptide selectivity where the second LC dimension is almost exclusively low pH RP. A variety of first dimensions can be used including RP, strong cation exchange (SCX), isoelectric focusing (IEF) and hydrophilic interaction

liquid chromatography (HILIC)[48]–[51]. The most commonly employed first dimension is SCX, which separates peptides based on charge and RP. As the majority of tryptic peptides have a charge of 2+ and 3+ there is limited scope in SCX to utilise the entire separation space. It has been previously shown during an SCX separation that the majority of peptides from a tryptic digestion elute in two main clusters. RP-RP separation has emerged as an effective alternative to SCX-RP[52]. In RP-RP the first dimension is carried out using a mobile phase with a pH ~ 10 as opposed to pH ~ 3. Tryptic peptides typically contain ionisable basic and acidic residues and peptide isoelectric constant (p*I*) is determined by the distribution of these residues. By dramatically changing the mobile phase pH between the first to second dimension of separation, peptide charge and therefore retention behaviour will be altered, providing the required degree of orthogonality. The use of RP-RP in the context of label-free quantitation was recently demonstrated by Patel and colleagues [53]. It was shown that RP-RP separation can increase protein identifications almost two-fold with only a minor increase in technical variation of quantitative results.

1.13 Protein Level Separation

Protein level separation, prior to proteolytic digestion is an alternative method for providing orthogonal separation to RP-LC. Protein level separation is most commonly carried out by SDS-PAGE where proteins are separated based on protein molecular weight in the presence of high concentrations of SDS. Sample lanes are then cut into a defined number of slices (anywhere from 4 - >20). Proteins contained within each slice are digested within the gel. Resultant peptides can be extracted from the gel with the use of acid and an organic solvent such as acetonitrile. SDS-PAGE separation provides a cheap and effective method for reducing sample complexity and is ideal for situations where sample amount is not limited. A major benefit of SDS-PAGE separation is the ability to use high concentrations of SDS during cell lysis/tissue homogenisation. The use of high SDS concentrations ensures the solubilisation of hydrophobic proteins, such as membrane spanning proteins, which can be under-represented when using in-solution digestion protocols. Where sample amount is limited, low recovery of peptides post in-gel digestion can pose problems with peptide recovery rates from gel slices being dependent on peptide physiochemical properties including size and hydrophobicity

1.14 Mass Spectrometry Analysis in Proteomic Workflows

As previously mentioned, mass spectrometry is now able to detect and quantify thousands of proteins in a single sample. There are a number of different mass spectrometry techniques that can be employed for the detection and quantification of a protein/s and indeed the ability to determine relative and absolute amount of proteins expressed by a cellular system at particular time points (TP) is an integral part of proteomics. Mass spectrometry is not, however, an inherently quantitative technique due to the ionisation efficiency of biomolecules being highly dependent on their amino acid sequence. Techniques for the quantification of proteins in a complex sample have developed rapidly over the past decade. There are now a large number of different quantification methods that can be employed for the analysis of a sample. The technique of choice depends on a number of factors including sample type and amount, time, instrument availability and cost. Broadly speaking the methods of quantification can be divided into two main categories label-based and label-free, **Figure 1.12**. Within these categories there are a number of common protocols each with benefits and limitations.

The earliest form of mass spectrometry-based quantitative proteomics was based on two-dimensional gel electrophoresis (2-DE). 2-DE is a gel-based separation technique, first proposed by P. O'Farrell in 1975,[54] where proteins are first separated based on their pI by isoelectric focusing. Following this, proteins are separated by their molecular weight at a 90 degree axis. Gels are then stained for protein quantification. Resultant spots correspond to a protein/protein group which can be characterised, after in-gel digestion, by MS analysis. Although this technique provides good protein resolving power it does have a number of limitations including the poor separation of hydrophobic, acidic and highly alkaline proteins [55], [56].

Over the past decade label-based quantitative techniques have become the chosen method for quantifying large numbers of proteins in a complex biological sample. Mann and co-workers' development of stable isotope labelling by amino acids in cell culture (SILAC) in 2002 was perhaps the biggest factor in popularising label-based quantification [57]. SILAC allows the direct comparison of biological samples within the same analytical run by the *in vivo* incorporation of 'heavy' amino acids into one of the biological samples being analysed. The method was first developed in cell culture experiments where one of the two cell cultures for comparison is grown in a culture media containing ^{13}C labelled lysine and/or ^{13}C labelled arginine. Samples can be combined post lysis and prior to

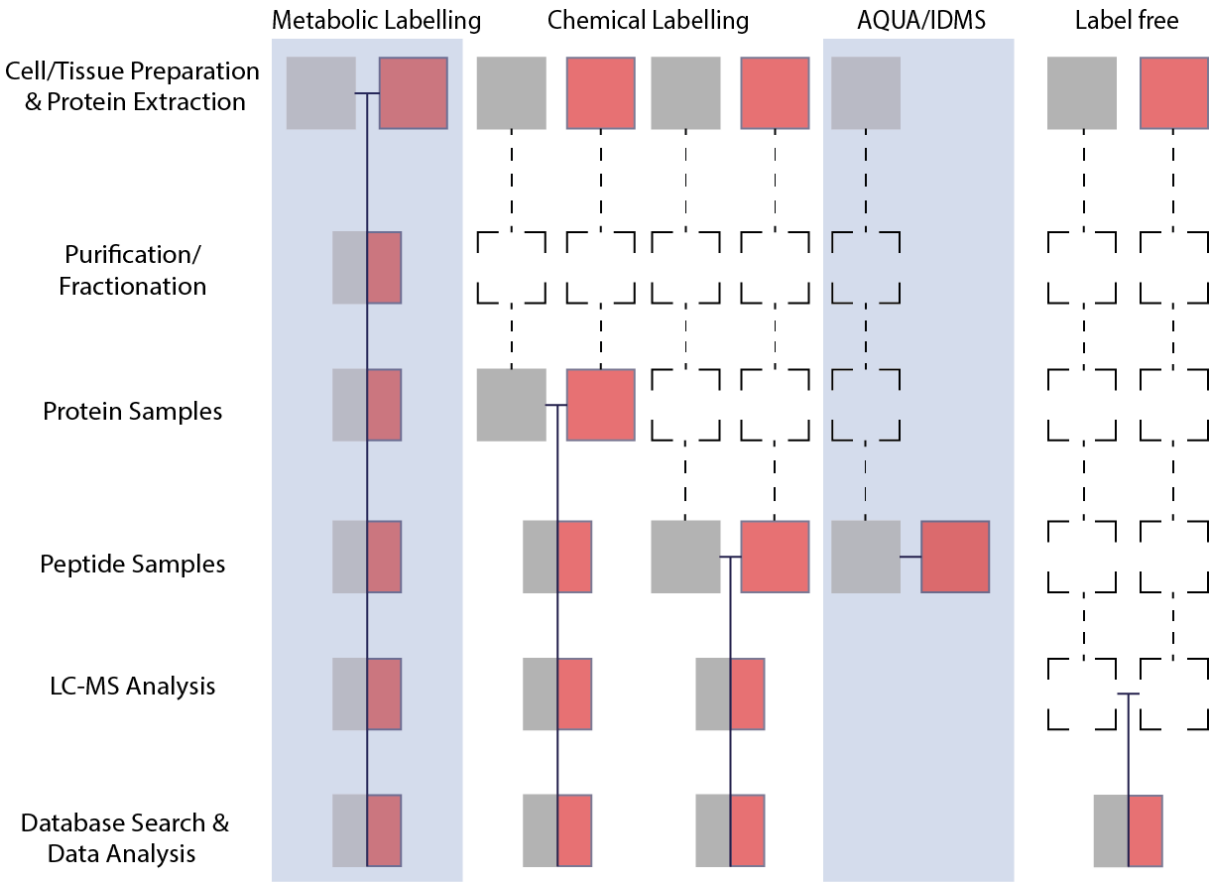


Figure 1.12 Overview of the four common quantitative workflows. Coloured boxes represent two experiment conditions. Horizontal lines that connect two boxes indicate when samples are combined. Empty boxes indicate that the samples are treated separately at each stage

digestion, minimising measurement uncertainties that can be introduced through pipetting errors and digestion efficiency differences. SILAC has been expanded to include a range of model organisms including mice, zebrafish and more recently *Drosophil.*[57]–[59]. The main limitations of SILAC protocols were previously the inability to study more than two biological conditions in an experimental workflow and that it was not applicable to human tissue samples. A protocol where SILAC labelled cells were used as a spike-in standard was developed to address these problems.[60] The concept was expanded to 'Super-SILAC' where a spike in standard is produced from a mixture of cell lines that differ in origin, stage and subtype [61]. This has been shown to be of particular benefit for cancer studies where tumour tissues show complex cell type diversity [62].

Other common label-based methods used for discovery proteomics are based upon isobaric labelling where quantitation is achieved at the MS/MS level through the detection of reporter ions. Two of the

most common methods are isobaric tags for relative and absolute quantification (iTRAQ) and tandem mass tags (TMT)[63], [64]. Both methods use the reactivity of primary amines to label lysine residues and/or the N-terminus of tryptic peptides with specific tags. In iTRAQ the tagging molecule contains three regions. The first region is the peptide reactive group which reacts with primary amines to form a covalent bond. The third group is a reporter ion which upon CID is fragmented. The reporter group can be changed in mass to vary the *m/z* at which it is detected. The second region is called the balancer group as this region's mass is altered in respect to the mass of the reporter group to ensure the combined tag mass remains consistent. Samples to be tested are labelled with a specific reporter, post-protein extraction and digestion. Post-labelling samples are pooled into a combined sample and analysed by LC-MS/MS. Relative quantities for each detected peptide can be attributed to each condition by comparing the intensities of the reporter ions detected in MS/MS scans.

Similar to SILAC, isobaric labelling methods allow the analysis of different samples within the same analytical run, leading to increased analytical precision. Indeed iTRAQ has been shown to offer higher precision quantitation than SILAC in some cases.[65] A major caveat with isobaric labelling methods is that one of the prerequisites for accurate quantitation is the isolation of single precursor peptides. The co-isolation of precursors can cause the measured peptide ratio to become skewed towards unity. This can be commonplace for complex samples and methods subsequently often under report fold changes, thereby increasing the false negative rate of quantitative results.

The use of isotopically labelled peptides as internal standards for protein quantification is still a gold standard method. The principles of isotope dilution mass spectrometry (IDMS) for relative and absolute quantification are well established [66], [67]. By using multiple isotopically labelled analogues of specific proteolytic peptides per target protein, quantitation precisions <10% can be easily achieved [68]. IDMS lends itself well to absolute quantification as exact amounts of the labelled standards can be spiked in for reference. Absolute quantification refers to the determination of protein concentrations, often in reference to total protein amounts, volumes or protein copy numbers per cell. IDMS however, is almost exclusively used for the validation of discovery proteomic results due to the cost of peptide internal standards and the fact that only a limited number of proteins can be quantified per analytical run.

A recently developed label-based method, QconCAT, aims to simplify the absolute quantification of large numbers of proteins through the synthesis of artificial proteins that are concatamers of surrogate peptides for the protein/s to be quantified.[69] The labelled protein or QconCat, which has been expressed, purified and quantified, are spiked into samples prior to digestion to act as an internal standard in a similar manner to IDMS. Endoproteolytic digestion of the sample releases surrogate peptides from the QconCat at a 1:1 ratio which can then be used for absolute quantification. The protocol represents a theoretically simplified method for quantifying large numbers of proteins in comparison to traditional IDMS methods. In addition, having the labelled peptides spiked as a pseudo-protein should help reduce the potential for the modification or degradation of standards during the digestion protocol. In practise, however, a number of challenging steps must be addressed for this protocol to work effectively including efficient expression and complete digestion.

Early label-free quantitation methods were primarily based on peptide or spectral counting methods where the number of fragment spectra assigned to a given protein is used as a proxy for protein abundance [70], [71]. The method is based on the observation that the more abundant a protein is the more MS/MS spectra are collected for peptids derived from the protein in question, **Figure 1.13**. This approach is often referred to as a semi-quantitative technique as it allows differential protein expression to be detected but the measurement of exact fold changes or absolute amounts are not robust [72]. The spectral counting method, although easy to implement, has never been fully adopted as a routine quantitative method. Spectral counting assumes a linear response for every protein however it is known that spectrum counts are peptide specific as changes in retention time, peak width and fragmentation efficiency can vary. Despite this a number of methods have been devised for calculating absolute protein expression levels from spectral counting values. emPAI, one of the higher profile methods takes into account the number of possible tryptic peptides that could be detected for each protein [73]. The method, however, still produces a large amount of error and even with more advanced computational methods, absolute amounts can still only be determined with an order of magnitude error.

An alternative label-free quantitative method that has seen a recent surge in popularity is based upon using precursor (or fragment) ion abundance to calculate protein concentration [74]–[76]. Relative peptide abundances can be determined by either calculating the area under the curve (AUC) or the

apex of peptide chromatographic peak, **Figure 1.13**. The method is based around the observed property of ESI efficiency being directly proportional to analyte concentration. Peptide intensity, therefore, correlates extremely well with its abundance. The relative levels of a protein, therefore, can be easily compared from sample to sample by the summation of peptide ion intensities derived from their parent protein. The improvement of instrument and chromatographic resolving power has continued to improve label-free quantitation through the reduction of potentially co-eluting and isobaric peptides which can interfere with quantitation. It was however the development of nESI that had the biggest contribution to ion-based label-free quantitation. As previously mentioned, nESI has a much larger tolerance for ion suppression effects which is of particular importance when samples to be compared are analysed in separate LC-MS/MS runs. The use of nESI allows the comparison of samples whose background may differ slightly in constituents due to the minimisation of potential matrix effects on peptide ionisation efficiencies.

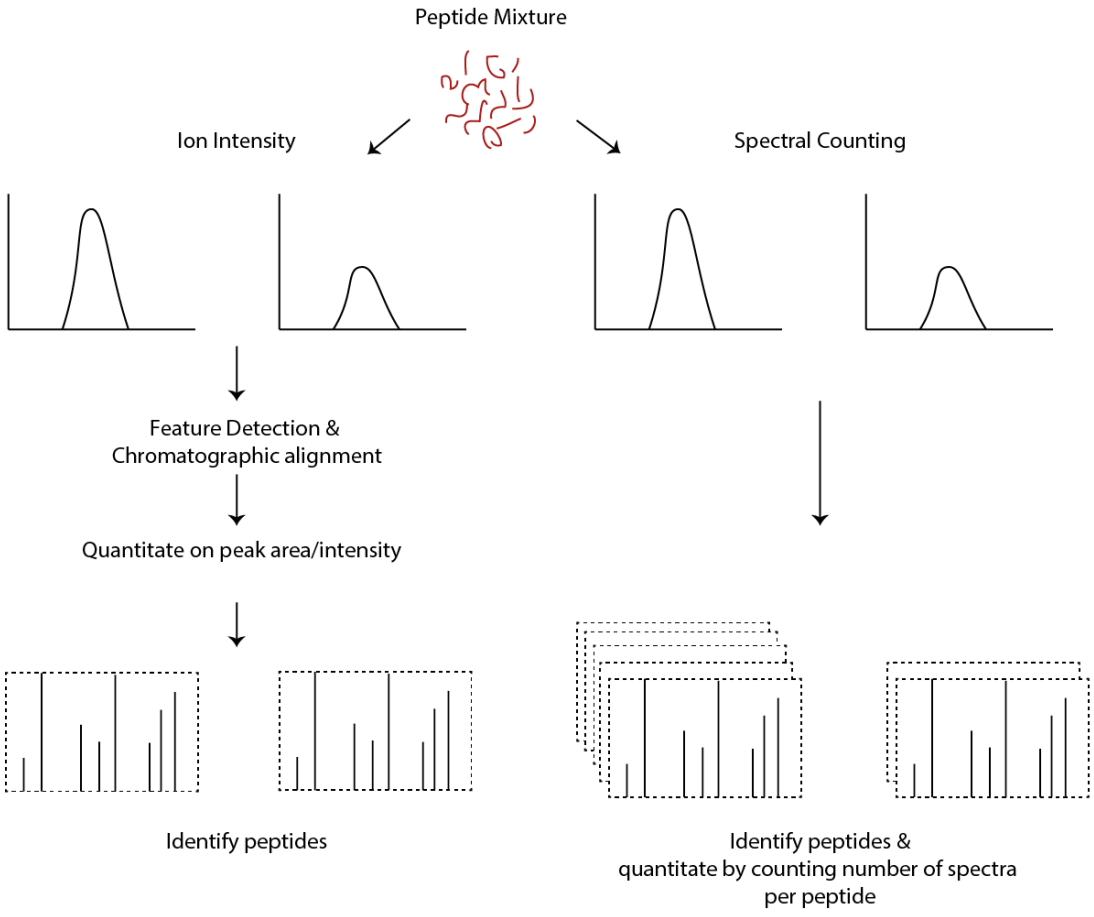


Figure 1.13 Overview of two label-free quantification approaches; Ion intensity-based and spectral counting

The use of ion-based, label-free relative quantitation is now commonplace. This is well demonstrated by the adoption of label-free quantitative functionality within the MaxQuant software developed by the Mann lab, software which was initially developed to perform SILAC data analysis [77]. Recent approaches have developed label-free techniques to allow the estimation of absolute protein amounts with good success. The approach developed by Selbach and co-workers, intensity-based absolute-quantification method (iBAQ), uses the sum of extracted ion intensities of all identified peptides per protein and normalises this value to the number of peptides theoretically produced by enzymatic digestion [78]. These values can be converted to copy numbers per cell if the number of cells used in an experiment is known. This value is easily quantifiable when analysing cell lines but is far more challenging when working with tissues. Mann and colleagues have utilised the depth of proteome coverage possible with current mass spectrometry technologies, to estimate absolute values without the need for cell counting [79]. Using their method total sample protein content can be assumed to be represented by the sum of all peptide ion intensities detected from a sample. The ratio of MS signal for a particular protein with respect to total signal intensity is therefore considered a proxy for its absolute amount per total protein amount. Copy numbers per cell can be derived from these ratios by dividing the values by protein molecular weight, multiplying by the Avogadro constant and the average protein content of a single cell. This technique, referred to as total protein approach (TPA) can be theoretically applied to sample where total protein can be determined by an assay such as the Bradford. The same group more recently has taken their approach forward and are now able to calculate absolute amounts based on MS data and without the need for any other sample assay information. In what they term the 'Proteomic Ruler', they demonstrate that the MS signal attributable to histones can be used to accurately determine cell numbers based on the proportionality of total histone concentration to DNA amount [80].

An alternative methodology for determining absolute protein amounts (in addition to relative quantification) was first published by Geromanos and co-workers and is commonly referred to as Hi3 quantification [81]. Hi3 utilises the unexpected discovery that the MS signal response from the three highest ionising peptides derived from each protein correlates ($R^2 > 0.95$) to the original protein concentration. Results using this method typically have a technical CV of <10-15%. To convert intensities into absolute values an internal standard, usually a single protein digest, can be spiked into the sample of interest at a known amount, an approach analogous to the IDMS/AQUA approach. The

use of an internal standard has the additional benefit of controlling for run to run loading amount and instrument sensitivity variability.

Although label-free approaches still do not provide the precision and accuracy of some label-based techniques, data quality is now at a point that biological questions can be answered without the need of expensive labelling reagents. Label-free quantitation is applicable to any biological sample that can be analysed by mass spectrometry and the numbers of conditions that can be directly compared are theoretically limitless. An additional benefit of the label-free approach is that it can be applied to any dataset post analysis and can even be used to compare datasets that were produced by different labs. A caveat of this is that label-free quantitation relies upon sample preparation being identical for each sample being analysed. Normalisation techniques or the use of relative changes can address this issue to a certain extent. In a recent seminal paper that presented the first mass spectrometry derived draft of the human proteome, label-free quantitation was used to estimate absolute protein abundances for 27 human tissues and body fluids.[82] The data used were derived from a number of different research labs with some obtained from public data repositories. The authors found that data were easily comparable without the need for extensive normalisation. Interestingly in the same publication, a comparison of Hi3 and iBAQ showed little difference in measurement error apart from Hi3 showing slightly higher error when the number of detected peptides per protein is low (<5), a characteristic that has been observed previously for Hi3 [83], [84].

Quantitation methodology is not the only technical consideration prior to beginning a quantitative proteomics experiment. One of the benefits of MS is the large number of acquisition modes that are available with each one having their own benefits and drawbacks. Depending on the aims of an experiment a particular mode may an obvious choice. One example is that the validation of previous quantitative results would typically require a targeted mode where reducing data dimensionality will reduce measurement errors. Discovery proteomic experiments, by their very nature, use an untargeted approach to quantify proteomes in an unbiased approach. Within this category, however, there are two distinct sub categories that define the acquisition approach taken.

1.15 Data Dependent Acquisition (DDA)

Data dependent acquisition (DDA) is the traditional “bottom up” approach for analysing a complex protein mixture. A DDA experiment consists of a series of precursor selections, isolations and fragmentations. A cycle typically consists of a survey MS scan to determine the *m/z* and relevant intensities of precursors currently eluting from the chromatographic column. Precursor ions are isolated serially for fragmentation, for either a set time period or until an ion current threshold is reached, **Figure 1.14**. The order in which they are isolated is almost exclusively based on intensity, with the most common approach being to target peptides for MS/MS from the most abundant to the least. Depending on instrumentation speed, the maximum number of peptides that can be targeted following a survey is currently ~ 30 and this number can be set by the user. Once the maximum peptide number is reached, or all detectable peptide ions have been sequenced, the cycle starts again with another survey scan.

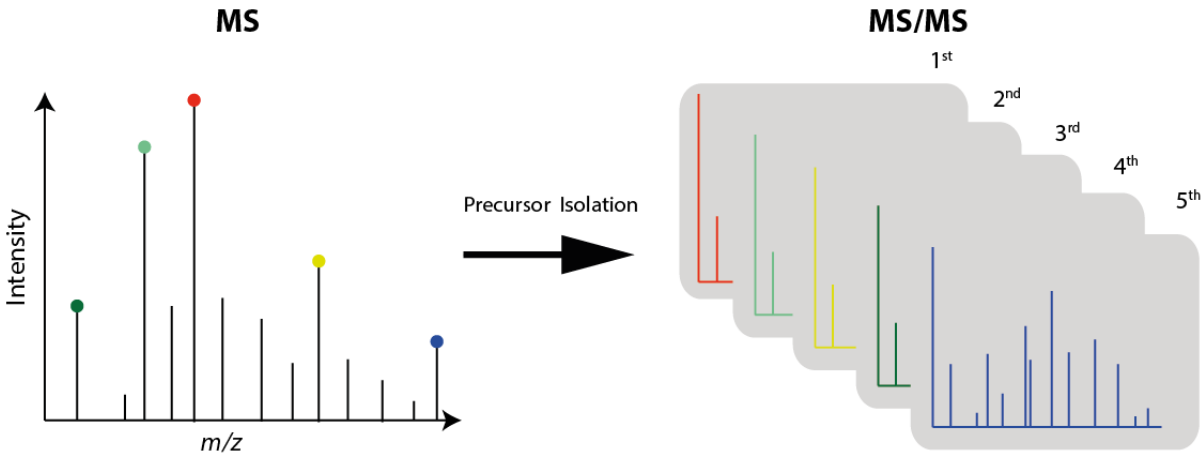


Figure 1.14 Data dependent methodology. MS scan identifies 5 most abundant precursor ions. Precursors are sequentially quadrupole isolated and fragmented in order of intensity

The intensity based selection of DDA is one of the contributing factors to MS’ bias towards the identification of abundant proteins, as these will inherently produce more abundant tryptic peptides. A further problem is its contribution to the stochastic nature of peptide identifications. Technical replicates, although identifying a large number of common peptides will also identify peptides that are not observed in one of the replicates, despite the analysis of a sample of identical constituents.[85] In

addition to ESI fluctuations, this is thought to occur due to run to run variability in LC and MS time alignment. If the start of the DDA series is different from run to run, in reference to LC sample injection, peptides may or may not be at their chromatographic apex when a survey scan is undertaken. This will influence the selection of peptides for fragmentation and result in the loss/gain of peptides that are near the threshold for selection.

1.16 Data Independent Acquisition (DIA)

An alternative acquisition mode that is beginning to see more widespread adoption is data independent acquisition (DIA) [86]. DIA encompasses a number of different strategies that pertain to remove the need for precursor selection for identification. The motivation behind the removal of precursor selection, during MS/MS analysis, is to achieve a more unbiased analytical approach. One such approach developed in the Aebersold lab is “SWATH MS” which takes its name from the sequential isolation windows (or SWATHs) that are acquired during an acquisition [87], [88]. During a SWATH analysis, precursor isolation windows, each 26 Da wide and overlapping by 1 Da, are used to scan across a pre-determined mass range (nominally 400-1200 m/z). MS/MS data is acquired for all precursors within the isolation window. Identification of peptides is achieved by matching MS/MS data with a spectral library that is usually produced by prior in depth DDA analysis of a sample representative of those being analysed. Protein quantification is achieved at the MS/MS level through the combination of fragment ion chromatograms that are derived from the same protein. Each cycle of SWATHs across the mass range takes 3.2 sec, which based on a chromatographic peak of ~30 s (if using HPLC) provides 9 points across a peak and is adequate for quantitation. The method has been shown to provide a quantitative CV of < 10 % for technical replicates which compares well with more targeted approaches.

The reliance of SWATH on spectral libraries for identification,[89] which are in the main created with the use of a DDA approach, means that SWATH is more of a ‘semi-DIA’ method. The concept of MS^E, which completely removes the need for DDA derived data, was introduced by J.C. Silva and colleagues [83], [90]. During an MS^E, acquisition MS and MS/MS data are recorded concurrently by altering the energy within the collision cell, **Figure 1.15**. No pre-selection of precursors is made and the quadrupole is used only to transfer ions within the expected *m/z* range that tryptic peptides are

predicted to fall within, usually 300-2000 *m/z*. Post data acquisition, fragment ions are assigned to their respective precursor ions based primarily on chromatographic time correlation. This process is refined during the database search to include information of the physiochemical properties of detected peptides when undergoing collision induced dissociation. The assignment follows an iterative depletion process where single proteins are identified sequentially and the associated peptides and fragments are removed from the dataset that is being mined. The result is a protein-centric search which provides high protein coverage. The concept of MS^E builds on the work of Clemmer and co-workers who, on their *in-house* built instrument collected MS and MS/MS data across the entire mass range [91]. This early incarnation of MS^E was named 'parallel fragmentation approach' for obvious reasons.

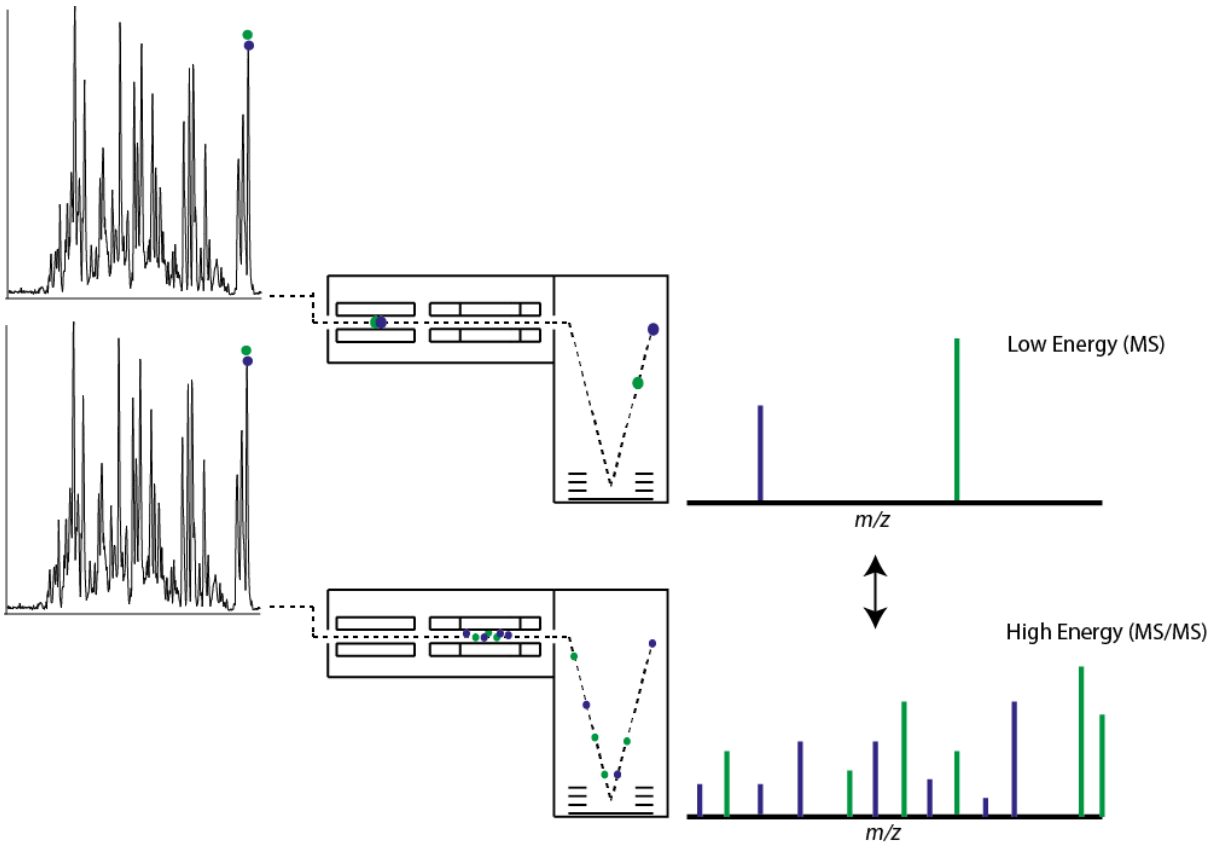


Figure 1.15 MS^E approach to data independent acquisition. As peptides elute the quadrupole is set to transmit all peptides ions. Energy within the collision cell is continuously alternated between low energy and a high energy ramp across the entire chromatographic run, allowing simultaneous acquisition of MS and MS/MS data

Interestingly Clemmer and co-workers were not the only group, at this point in time, working on the idea of multiplexed peptide MS/MS measurements, with the ultimate aim of higher-throughout peptide identification rates. Work published by the group of R. D. Smith successful demonstrated that the use of a high resolution instrument, in this case tandem FTICR MS, allowed the simultaneous fragmentation of polypeptides and their successful identification [92]. Some years later Yates and colleagues presented data showing the benefits of the data independent acquisition approach, using a method, somewhat analogous, to the now popular SWATH method [93]. By sequentially isolating and fragmenting across precursor windows of 10 Da, within an ion trap mass analyser, it was shown that signal-to-noise was consistently increased by three-to fivefold in comparison to a traditional DDA methodology. This concept was further developed, through the incorporation of high resolution instrumentation, namely an orbitrap mass analyser, by the group of D.R. Goodlett [94]. In their method, precursor acquisition independent from ion current (PAcIFIC), by scanning across very small isolation windows (typically 2.5 Da), whilst collecting MS/MS data for all ions, it was shown that hundreds of proteins could be detected across 8 orders of magnitude, a range that was not achievable by traditional DDA.

With no pre-selection of precursors, MS^E aims to collect MS and MS/MS data for all peptides across the whole mass range as they elute from the chromatographic column. For all peptides where sufficient quality data is acquired a successful identification should be made, opening up the potential for peptide identification rates that surpass that of DDA runs. Whereas DDA methods are limited by scan speeds, MS^E and other such DIA methods are theoretically limited only by system peak capacity. Another major benefit is a close to 100% duty cycle that increases the number of points across a peak that can be used for quantitation. Quantitation by MS^E is on MS level data where, per scan, the instrument spends half its time acquiring. If a 1 sec scan speed is considered, typical for an MS^E experiment, then 0.5 sec will be spent acquiring MS data and the remaining 0.5 sec spent acquiring MS/MS data. Including the inter scan delay of ~0.2 sec, MS^E records 25 data points across a 30 sec chromatographic peak. With the high duty cycle that MS^E provides comes its applicability to UPLC chromatography where due to the high level of resolving power peptides can elute with a peak width at half height (PWHH) of < 10 sec. To achieve sufficient data points for high quality quantitation (15-20 ideally) across such high resolution peaks a very high duty cycle is required. Indeed MS^E coupled to UPLC has been demonstrated to provide good technical reproducibility, typically < 15 %[81].

Although the theory behind MS^E does offer the potential for improved peptide identification rates in comparison to DDA, initial proteome coverages achieved were not as high as those achieved on high resolution instruments. MS^E is only available on Q-ToF instruments developed by Waters Ltd, whilst the majority of high proteome coverage datasets were achieved with the use of Orbitrap technology, so differences between instrument sensitivity and resolution cannot be discounted as contributing factors. The development of the Synapt G2, followed more recently by the Synapt G2-S, introduced ion mobility based peptide separation with QToF mass spectrometry and went some way towards realising the potential of MS^E[94], [95].

High resolution mass spectrometry allows the separation of potentially isobaric co-eluting peptides by small mass differences, however the development of ion mobility coupled to mass spectrometry has provided an alternative 'on the fly' method for increasing proteome coverage. Until the development of the commercially available Synapt instruments, ion mobility mass spectrometry (IMMS) was only available to a small number of labs with custom made instruments [96]–[98]. Ion mobility separates gas phase ions prior to the mass analyser, based upon their mass, charge and shape (rotationally averaged, or collision cross section), **Figure 1.15**. Synapt instruments use traveling wave ion mobility separation (TWIMS) to achieve this separation. In TWIMS ions are accelerated through the mobility chamber by a series of stacked ring ions guides. The mobility chamber is kept at low pressure, relative to the high vacuum region of a mass spectrometer, by the introduction of nitrogen gas. Pulsed ions accelerated through the mobility cell undergo collisions with nitrogen gas molecules with the number of collisions experienced by a molecule being proportional to its size. The result is that smaller molecules traverse the mobility cell quicker than large molecules. TWIMS has until recently been primarily used to probe the conformations of proteins and protein complexes due to the earliest Synapt platform, Synapt HDMS, not being suitable for peptide analysis due to saturation effects [99].

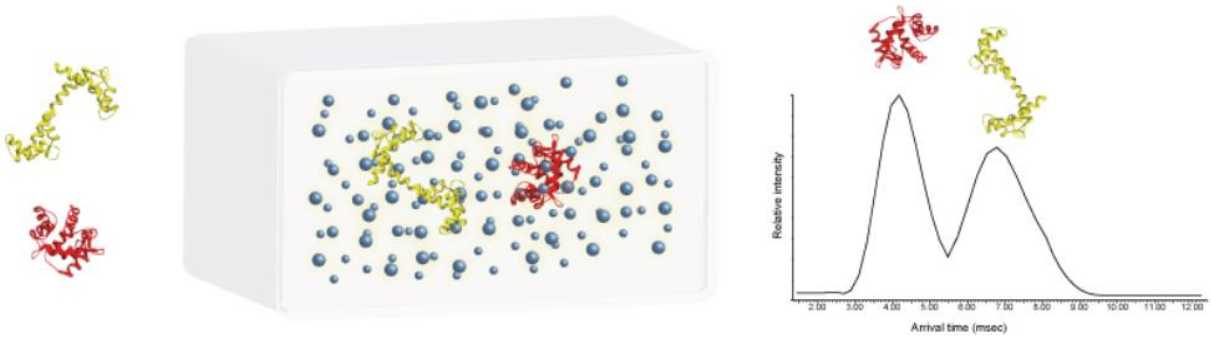


Figure 1.15 Principle of ion mobility separation. Two protein ions of the same *m/z* are separated based upon their collision cross section. Here, the more extended protein (yellow) traverses the mobility cell more slowly as it experiences more collisions with the buffer gas than the more compact protein (red)

The Synapt G2 saw the use of an ADC detector in place of the previous TDC to reduce the problem of saturation effects, although the effects are still present for highly abundant ions [100]. The introduction of ion mobility adds an extra dimension of separation which is orthogonal to RP and occurs on a timescale that fits between LC and MS, **Figure 1.16**. HDMS^E, as it is referred to, uses both chromatographic retention time and gas phase mobility to align precursors with their respective fragments. The additional mode of separation during HDMS^E provides protein and peptide identification rates up to 60 % higher than achieved using MS^E on the same instrument. An additional benefit of HDMS^E acquisitions are increased quantitation precision for low abundance peptides and increased peptide detection dynamic range. This along with the saturation effects which are apparent on the Synapt HDMS are partly a result of the compression of ions in space during ion mobility separation. This process is analogous to the enhanced sensitivity provided by chromatographic separation. A further reason for increased quantitative precision and dynamic range during HDMS^E is a reduction in co-detected species, including both isobaric peptides and contaminants, through their cross sectional based deconvolution.

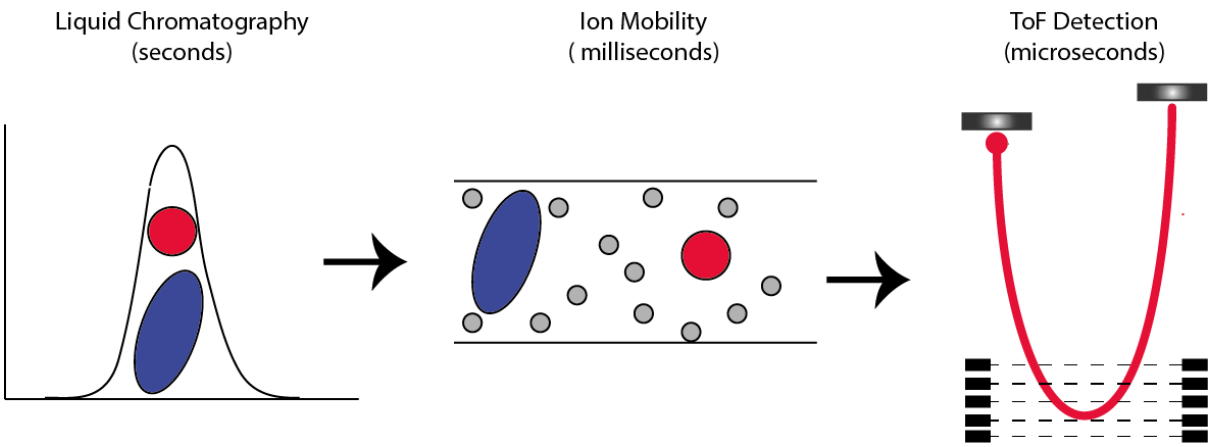


Figure 1.136 Complementarity of ion mobility as an orthogonal separation method to liquid chromatography mass spectrometry.

Even with the addition of IMS, $HDMS^E$ is still not able to provide the identification rates and the proteome coverages at a level similar to those of DDA-based workflows. Tenzer and colleagues were able to demonstrate that a major limiting factor of both the MS^E and $HDMS^E$ approach was inefficient precursor ion fragmentation [101]. During MS^E a collision energy ramp is conducted during the elevated energy scan where in $HDMS^E$ mode a fixed collision energy is used during each individual IMS cycle and across the entire *m/z* range. This is in contrast to DDA-based methods where collision energies can be determined based on precursor *m/z*, an approach that is not possible in DIA-based methods. Tenzer and co-workers were instead able to use ion mobility profiles to improve fragmentation efficiency. In $UDMS^E$ the collision energy is ramped with reference to the ion mobility profile during each IM cycle. Peptides that traverse the cell quickly and are therefore more likely to be small and/or highly charged are subjected to lower collision energy whilst peptides that arrive later in the IM cycle are fragmented with higher collision energy. Using the optimised collision energy profiles, fragmentation efficiency was significantly improved (105% increase) in comparison to both MS^E and $HDMS^E$. Proteome coverage for a HeLa cell lysate was increased by 47 % in comparison to $HDMS^E$ and importantly it was shown that $UDMS^E$ had a larger peptide identification rate than is theoretically possible on current generation Orbitrap instruments in DDA-mode.

1.17 Database Searching

The emergence of gene sequencing technology has facilitated the complete nucleotide sequencing of commonly used cell lines and organisms. Genome wide sequences for each organism can subsequently be translated to form a predicted protein sequence database. The availability of predicted protein sequence databases has been a significant factor in the rapid popularisation of mass spectrometry for proteomic analyses.

The interpretation of mass spectrometry proteomic data relies upon the use of enzymes with predictable cleavage sites and the well characterised fragmentation of peptides by CID/ETD. Along with the sample origin, which is used for the selection of the correct protein sequence database, these factors allow the prediction of all possible peptide and peptide fragment masses that could be produced from a theoretical mass spectrometry experiment. Peaks extracted from the raw mass spectrometry data are then searched against the *in-silco* data in order to identify peptides from the sample in question. The use of an enzyme with well-defined cleavage sites means that the identification of proteins can be inferred by the direct identification of peptides.

Due to the vast nature of possibilities that are produced from such *in-silco* predictions the possibility of false discoveries is a well-documented problem in the proteomics field. There are a number of statistical algorithms that have been developed to allow the proportion of incorrect identifications to be controlled [102]–[104]. The most common approach is to measure the false discovery rate by using a target decoy database search strategy [105]. This approach creates a composite database by concatenating the target protein sequence database with a decoy database. The decoy is created by either reversing or randomising the protein sequences from the target. By doing this the overall residue composition of the target database is retained but the two databases contain practically no common peptide sequences. Two assumptions are made in this approach which was validated by J.E Elias and S.P Gygi in an early methods paper documenting the approach [106]. The first assumption is that sequences within the target and decoy database do not overlap as if they did this would cause an erroneously high FDR estimation. Elias and Gygi found that for peptides over 7 residues there were only 0.02% common sequences between the two databases. The second is that false positives are equally likely in both the target and decoy. This was again validated by using an *in-silco* approach to show that distribution of considered peptides for two common database search programs was

equal for both database types. Interestingly, the agreement between the target and decoy only holds true when the search space is large enough. Although it is commonly accepted that the use of very small databases will incorrectly measure the FDR for this reason, the effect of using high mass accuracy searches on the FDR has only been assessed by a small number of groups. This is of particular importance now that the vast majority of proteomic experiments are performed with instruments that offer mass accuracies < 5ppm. This high mass accuracy is often used to reduce database search space which has been shown to have deleterious effects on FDR determination [107].

Database searches of mass spectrometry data are performed by dedicated search engines of which there are many [108]–[112]. Although each search engine essentially follows the same basic premise there are important differences which are mostly how the different software packages determine the confidence of peptide identification. Four common search engines are Sequest developed by the Yates lab, Mascot, developed by Matrix Science, Andromeda, developed by the Mann lab and ProteinLynx Global Server (PLGS), developed by Waters Ltd. Other popular search engines include OMSSA, X!Tandem and CRUX.

Sequest, first presented in 1994, is one of the original search engines used to automatically identify MS/MS spectra based on a protein sequence database [113]. Briefly for each MS/MS spectra precursor mass is used to generate a list of *in-silco* peptides that have a mass close to the observed mass. For each peptide theoretical fragment spectra are compared with the raw data and the candidate peptide with the closest matching spectrum is reported as the best identification.

Mascot was one of the earlier search engines and follows a similar approach to Sequest. Mascot has been adopted by many as it is designed to process a wide range of mass spectrometry data derived from various different instruments [109]. Mascot uses a probabilistic scoring algorithm for protein identification that is based upon the molecular weight search (MOWSE) algorithm. The algorithm calculates the probability that an observed match between experimental data and the protein sequence database is by chance. The best match is therefore the match with the lowest probability. Whether the best match is significant depends on the size and database and Mascot provides a significance threshold which is based upon a p value < 0.05. Mascot is a search engine only and is not designed to process raw data into the MS and MS/MS peak lists required as input for the search.

Proprietary software is available for all main instruments to generate the mascot generic format (mgf) from raw data.

MaxQuant is a more recently developed search engine that has been designed to process high resolution data, produced primarily from Orbitrap mass analysers. In its early form, MaxQuant performed feature detection and SILAC quantification but relied upon Mascot to determine peptide candidates for MS/MS spectra [114]. MaxQuant now has its own dedicated search engine, Andromeda [110]. MaxQuant uses the high resolution provided by FT-based mass analysers to identify mass features at the MS level. The centroid mass of each peak is then determined by fitting a Gaussian peak shape to the three most central data points. The peak is then represented in 3D by including the retention time elution profile. Mass accuracy is improved by using SILAC charge state pairs, assigned prior peptide identification, to recalibrate the data post acquisition. Mass accuracies in the ppb region are possible with this approach. Andromeda uses a probabilistic scoring function which has been shown to perform as well as Mascot and is suitable for scoring post-translational modifications such as phosphorylation.

ProteinLynx Global Server software has been developed specifically to processes MS^E data along with DDA data [90], [108]. The program subsequently has novel processing steps that are not required for DDA derived data. As mentioned because of the method of acquisition during MS^E fragment ions must be assigned to their parent precursors post acquisition, which is currently only possible with PLGS. Prior to fragment assignment two algorithms are used to process raw data. Apex3D subtracts noise before integrating ion current signal across the entire chromatographic profile. The ion outputs from this stage are converted into exact mass and retention time (EMRTs) by the Pep3D algorithm which de-isotopes and charge state reduces all common peptides. It is at this point that fragments are initially assigned to their precursor counterparts based upon their elution and mobility ($HDMS^E$) profiles. This process is not highly discriminatory and will associate fragment ions with a number of precursors. The Ion Accounting algorithm performs the database search in a protein centric manner. Three stages of searches are performed starting with the identification of the most confident peptides and the removal of their EMRTs and associated fragments. This stage or “Pass1” continues until the protein FDR, as calculated by a decoy database, is reached. The second stage, Pass2, searches the remaining data for peptides derived from proteins identified from the previous

stage. Here peptides with missed cleavages, modifications and those derived from in-source fragmentation can be identified. Again the EMRTs of identified peptides are removed from the search space. The final search allows fragment ion intensity to be higher than its precursor, an attribute indicative of the in-source fragmentation of a highly labile peptide.

1.18 Spectral Libraries

A spectral library consists of a database of well characterised fragmentation patterns for the various compounds that are to be analysed, which in the case of proteomics are almost exclusively tryptic peptides. The identification of an unknown compound or peptide is made by successfully matching an unknown spectrum to a spectrum contained within the library. In small molecule analysis libraries are compiled by collecting fragmentation patterns from the analysis of known compounds [115], [116] whereas in proteomics such libraries can be produced from the results of extensive DDA analysis of a proteomic sample followed by database searching [117]. Spectral libraries are widely used in the analysis of small molecules where fragmentation is much harder to predict. Due to the predictable nature of peptide fragmentation, which facilitates database searching methods, the use of spectral libraries is far less popular. Spectral libraries do have their benefits in comparison to more traditional database searches and the development of DIA approaches, in particular SWATH, has seen their use in recent times increase [89]. Additionally the work of Thalassinos and colleagues has demonstrated its applicability to MS^E and $HSMS^E$ data [118]. By developing identification metrics based on precursor, product ion exact mass, peptide fragmentation efficiency, peptide retention time and mobility they were able to show that the use of a well validated spectral library can be successfully used to process DIA data.

Although traditional searching methods are well suited to the identification of new sections of proteomes that may have not been previous observed, spectral library searching is more efficient at providing in-depth coverage of proteome sections previous characterised. Due to the reduced search space the use of spectral libraries also offers faster searching speeds, a benefit that is becoming more important as the data volume produced by latest generation instruments increases. As DIA and targeted data validation becomes more widespread it would be expected that the use of spectral libraries will increase dramatically.

1.19 Protein Aggregation and Disease

Proper folding of proteins is essential for correct cellular function. Nascent polypeptide chain folding is dependent on both the intrinsic properties governed by the amino acid sequence and influences from the cellular environment. Without correct tertiary structures proteins are unable to perform their respective functions and as such the cell has a large number of quality control mechanisms that act to promote proper protein folding and identify misfolding [119], [120]. The difficulty of maintaining proteostasis cannot be understated, with thousands of different proteins, over a large concentration dynamic range, to be persevered in various environmental and metabolic conditions. Additionally the folding mechanisms of proteins mean that their native structures tend to be marginally stable at physiological conditions [121]. Under certain circumstances protein misfolding can overload the quality control mechanisms and cause misfolded protein load to increase. Alongside the high total protein concentration within cells (~300 g/l) such an increase leads to a high propensity for misfolded proteins to aggregate within or in the area surrounding cells.

The aetiology of diseases such as Alzheimer's disease (AD) and systemic amyloidosis have long been linked with the aggregation of proteins [122], [123]. More recently a far larger range of diseases have been identified as being caused by or have been correlated with protein misfolding and aggregation.[124] Such diseases are commonly associated with mutations that destabilise the native structure of the protein or prevent a protein ever reaching its native fold. It is however interesting to note that around 30% of proteins in higher eukaryotes have intrinsically disordered regions giving rise to a metastable nature increasing their propensity to aggregate [125]. The archetypal example is α-synuclein which is neuropathologically linked with Parkinson's disease (PD) where it forms well characterised Lewy bodies within nerve cells [126]. Indeed AD and PD are stereotypical misfolding diseases in that they both affect the central nervous system. In both diseases the biggest risk factor, disregarding any of the more severe mutations, is age. An ageing population has led to a dramatic increase in misfolding disease prevalence in western countries [127]. It is probable that this risk factor is linked to the decline in quality control mechanisms, such as chaperoning systems, upon senescence. It has been demonstrated by transcriptional profiling of C*aenorhabditis elegans* that a wide range of heat shock proteins display age-related mRNA level decline [128].

Pathological symptoms as a result of protein misfolding generally fit into two broad categories. A loss of function can be caused by a reduction in native protein availability due to only a percentage of synthesised protein adopting the correct fold. Gain of function symptoms conversely are thought to be caused by the toxic properties of oligomers and extended aggregates that arise from the exposure of hydrophobic regions upon misfolding. An alternative mechanism that leads to gain of function symptoms doesn't directly implicate the toxicity of oligomers and aggregates. Instead the overloading of available degradation and molecular chaperoning systems leads to a collapse of proteostasis creating a self-propagating cycle leading to eventual cell death. A key step towards developing treatments for misfolding diseases is to dissect the mechanisms by which pathological symptoms are induced. This is most accessible through the use of cell or animal models and indeed a number of important developments for various misfolding diseases have been enabled by such models.

1.20 Aims and Objectives

Although the aetiology of misfolding diseases are on the whole understood, the downstream events that result in clinical symptoms are far more complex. A major question in many misfolding diseases is how misfolded aggregates or oligomers induce damage to surrounding cells. Although the classic reductionist approach has its place in the study of misfolding diseases, the advancement and availability of mass spectrometry-based proteomics offers a more comprehensive and unbiased approach to characterising cellular responses to misfolding events.

The major objectives of my research were:

1. To develop a label-free quantitative proteomics workflow to identify differentially regulated proteins that are implicated in age-related vulnerability to Aβ42 toxicity in an adult-onset *D. melanogaster* model for Alzheimer's Disease
2. Use a data independent acquisition approach to conduct label-free quantitative proteomic time-course analysis of chronic Aβ toxicity in an adult onset *Drosophila* model for AD and compare with wild-type ageing. Use these data to identify candidate proteins and cellular responses that better explain the link between ageing and Alzheimer's disease

3. Develop a subcellular label-free quantitative proteomic workflow to characterise endoplasmic reticulum (ER) enriched protein responses to the expression of two misfolding variants of α_1-antitrypsin in comparison to the well folded, wild-type variant, in a Chinese hamster ovary cell line.

1.21 References

[1] J. Thomson, “Rays of Positive Electricity and Thier Application to Chemical Analysis,” *LongMans Green*, 1913.

[2] D. Cameron, A.E , Eggers, “An Ion Velocitron,” *Rev.Sci.Instrum*, vol. 19, 1948.

[3] F. McLafferty, “Mass spectrometric analysis aliphatic aldehydes,” *Appl.Spectrosc*, vol. 11, 1957.

[4] W. Paul and H. Steinwedel, “Ein Neues Massenspektrometer Ohne Magenfeld,” *Z.Naturforsh*, vol. 8a, pp. 448–450, 1953.

[5] W. Paul, H. Reinhard, and U. von Zahn, “Das elektrische massenfilter als massenspektrometer und isotopentrenner,” *Z.Phys*, vol. 152, no. 2, pp. 143–182, 1958.

[6] M. S. Munson and F. Field, “Chemical Ionization Mass Spectrometry.I.General Introduction,” *J. Am. Chem. Soc.*, vol. 88, pp. 2621–2630, 1966.

[7] A. Dempster, “A new method of positive ray analysis,” vol. 11, p. 316, 1918.

[8] M. Barber, R. S. Bordoli, R. D. Sedgwick, and A. N. Tyler, “Fast atom bombardment of solids (F.A.B.): a new ion source for mass spectrometry,” *J. Chem. Soc. Chem. Commun.*, p. 325, 1981.

[9] R. Macfarlane and D. Torgerson, “Californium-252 plasma desorption mass spectroscopy,” *Science*, vol. 191, pp. 920–925, 1976.

[10] A. Benninghoven, “Analysis of sub-monolayers on silver by secondary ion emission,” *Phys. Status Solidi*, vol. 34, p. K169, 1969.

[11] K. Tanaka, H. Waki, Y. Ido, S. Akita, and Y. Yoshida, “Protein and Polymer Analyses up to mlz 100 000 by Laser Ionization Time-of-flight Mass Spectrometry,” *Rapid Commun. Mass Spectrom.*, vol. 2, no. 8, pp. 151–153, 1988.

[12] M. Karas, D. Bachmann, U. Bahr, and F. Hillenkamp, “Matrix-assisted ultraviolet laser desorption of non-volatile compounds,” *Int. J. Mass Spectrom. Ion Process.*, vol. 78, pp. 53–68, 1987.

[13] J. Fenn, M. Mann, C. Meng, S. Wong, and C. Whitehouse, “Electrospray ionization for mass spectrometry of large biomolecules,” *Science*, vol. 246, pp. 64–71, 1989.

[14] L. L. M. Malcolm Dole, L. L. Marck, L. Hines, R. C. Mobley, L. D. Ferguson, and M. B. Alice, “Molecular Beams of Macroions,” *J Chem.Phys.*, vol. 49, no. 1968, pp. 2240–2249, 1968.

[15] J. V Iribarne and B. A. Thomson, “On the evaporation of small ions from charged droplets,” *J Chem.Phys.*, vol. 64, no. 1976, pp. 2287–2294, 1976.

[16] C. J. Hogan, J. A. Carroll, H. W. Rohrs, P. Biswas, and M. L. Gross, “Charge carrier field emission determines the number of charges on native state proteins in electrospray ionization,” *J. Am. Chem. Soc.*, vol. 130, pp. 6926–6927, 2008.

[17] M. Wilm and M. Mann, “Analytical properties of the nanoelectrospray ion source.,” *Anal. Chem.*, vol. 68, no. 1, pp. 1–8, 1996.

[18] M. S. Wilm and M. Mann, “Electrospray and Taylor-Cone theory, Dole’s beam of macromolecules at last?,” *Int. J. Mass Spectrom. Ion Process.*, vol. 136, pp. 167–180, 1994.

[19] W. Stephens, “A pulsed mass spectrometer with time dispersion,” *Phys. Rev.*, 1946.

[20] J. Snijder, R. J. Rose, D. Veesler, J. E. Johnson, and A. J. R. Heck, “Studying 18 MDa virus assemblies with native mass spectrometry,” *Angew. Chemie - Int. Ed.*, vol. 52, no. 400 V, pp. 4020–4023, 2013.

[21] R. J. Cotter, “Time-of-flight mass spectrometry for the structural analysis of biological molecules.,” *Anal. Chem.*, vol. 64, no. 21, p. 1027A–1039A, 1992.

[22] M. L. Vestal, P. Juhasz, and S. A Martin, “Delayed Extraction Matrix-assisted Laser Desorption Time-of-flight Mass Spectrometry,” *Rapid Commun. Mass Spectrom.*, vol. 9, no. July, pp. 1044–1050, 1995.

[23] B. Mamyrin, V. Karataev, D. Shmikk, and V. Zagulin, “The mass-reflectron, a new nonmagnetic time-of-flight mass spectrometer with high resolution,” *Sov Phys JETP*, vol. 37, no. 1, pp. 45–48, 1973.

[24] A. Makarov, “Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis,” *Anal. Chem.*, vol. 72, no. 6, pp. 1156–1162, 2000.

[25] P. Picotti and R. Aebersold, “Selected reaction monitoring–based proteomics: workflows, potential, pitfalls and future directions,” *Nat. Methods*, vol. 9, no. 6, pp. 555–566, 2012.

[26] P. Roepstorff and J. Fohlman, “Proposal for a common nomenclature for sequence ions in mass-sectra of peptides,” *Biomed. Mass Spectrom.*, vol. 11, no. 11, p. 601, 1984.

[27] K. Biemann, “Contributions of mass spectrometry to peptide and protein structure.,” *Biomed. Environ. Mass Spectrom.*, vol. 16, pp. 99–111, 1988.

[28] J. E. P. Syka, J. J. Coon, M. J. Schroeder, J. Shabanowitz, and D. F. Hunt, “Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, pp. 9528–9533, 2004.

[29] R. Zubarev, N. L. Kelleher, and F. W. McLafferty, “Electron capture dissociation of multiply charged protein cations. A Nonergodic Process", *J. Am. Chem. Soc*, vol. 120, no. 16, pp. 3265–3266, 1998.

[30] C. K. Frese, A. F. M. Altelaar, H. Van Den Toorn, D. Nolting, J. Griep-Raming, A. J. R. Heck, and S. Mohammed, “Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry,” *Anal. Chem.*, vol. 84, pp. 9668–9673, 2012.

[31] C. K. Frese, A. F. M. Altelaar, M. L. Hennrich, D. Nolting, M. Zeller, J. Griep-Raming, A. J. R. Heck, and S. Mohammed, “Improved Peptide Identification by Targeted Fragmentation Using CID, HCD and ETD on an LTQ-Orbitrap Velos,” *J. Proteome Res.*, vol. 10, pp. 2377–2388, 2011.

[32] C. Flangea, C. Schiopu, F. Capitan, C. Mosoarca, M. Manea, E. Sisu, and A. D. Zamfir, “Fully automated chip-based nanoelectrospray combined with electron transfer dissociation for high throughput top-down proteomics,” *Cent. Eur. J. Chem.*, vol. 11, no. 1, pp. 25–34, 2013.

[33] L. Fornelli, E. Damoc, P. M. Thomas, N. L. Kelleher, K. Aizikov, E. Denisov, A. Makarov, and Y. O. Tsybin, “Top-down analysis of monoclonal antibody IgG1 by electron transfer dissociation Orbitrap FTMS,” *Mol. Cell. Proteomics*, 2012.

[34] F. Lermyte, A. Konijnenberg, J. P. Williams, J. M. Brown, D. Valkenborg, and F. Sobott, “ETD allows for native surface mapping of a 150 kDa noncovalent complex on a commercial Q-TWIMS-TOF instrument,” *J. Am. Soc. Mass Spectrom.*, vol. 25, pp. 343–350, 2014.

[35] J. J. Coon, B. Ueberheide, J. E. P. Syka, D. D. Dryhurst, J. Ausio, J. Shabanowitz, and D. F. Hunt, “Protein identification using sequential ion/ion reactions and tandem mass spectrometry.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 27, pp. 9463–9468, 2005.

[36] R. Sachidanandam, D. Weissman, S. C. Schmidt, J. M. Kakol, L. D. Stein, G. Marth, S. Sherry, J. C. Mullikin, B. J. Mortimore, D. L. Willey, S. E. Hunt, C. G. Cole, P. C. Coggill, C. M. Rice, Z. Ning, J. Rogers, D. R. Bentley, P. Y. Kwok, E. R. Mardis, R. T. Yeh, B. Schultz, L. Cook, R. Davenport, M. Dante, L. Fulton, L. Hillier, R. H. Waterston, J. D. McPherson, B. Gilman, S. Schaffner, W. J. Van Etten, D. Reich, J. Higgins, M. J. Daly, B. Blumenstiel, J. Baldwin, N. Stange-Thomann, M. C. Zody, L. Linton, E. S. Lander, and D. Altshuler, “A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms.,” *Nature*, vol. 409, no. February, pp. 928–933, 2001.

[37] V. C. Wasinger, S. J. Cordwell, A. Cerpa-Poljak, J. X. Yan, A. A. Gooley, M. R. Wilkins, M. W. Duncan, R. Harris, K. L. Williams, and I. Humphery-Smith, “Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium.,” *Electrophoresis*, vol. 16, pp. 1090–1094, 1995.

[38] A. S. Hebert, A. L. Richards, D. J. Bailey, A. Ulbrich, E. E. Coughlin, M. S. Westphall, and J. J. Coon, “The One Hour Yeast Proteome,” *Mol. Cell. Proteomics*, vol. 13, pp. 339–347, 2013.

[39] A. M. Edwards, R. Isserlin, G. D. Bader, S. V Frye, T. M. Willson, and F. H. Yu, “Too many roads not taken.,” *Nature*, vol. 470, pp. 163–165, 2011.

[40] P. Picotti, B. Bodenmiller, and R. Aebersold, “Proteomics meets the scientific method.,” *Nat. Methods*, vol. 10, no. 1, pp. 24–7, 2013.

[41] A. Michalski, J. Cox, and M. Mann, “More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS.,” *J. Proteome Res.*, vol. 10, no. 4, pp. 1785–93, 2011.

[42] R. A. Zubarev, “The challenge of the proteome dynamic range and its implications for in-depth proteomics.,” *Proteomics*, vol. 13, no. 5, pp. 723–6, 2013.

[43] Y. Shen, R. Zhao, S. J. Berger, G. A. Anderson, N. Rodriguez, and R. D. Smith, “High-efficiency nanoscale liquid chromatography coupled on-line with mass spectrometry using nanoelectrospray ionization for proteomics,” *Anal. Chem.*, vol. 74, no. 16, pp. 4235–4249, 2002.

[44] H. Liu, J. W. Finch, M. J. Lavallee, R. A. Collamati, C. C. Benevides, and J. C. Gebler, “Effects of column length, particle size, gradient length and flow rate on peak capacity of nano-scale liquid chromatography for peptide separations,” *J. Chromatogr. A*, vol. 1147, pp. 30–36, 2007.

[45] X. Wang, W. E. Barber, and P. W. Carr, “A practical approach to maximizing peak capacity by using long columns packed with pellicular stationary phases for proteomic research,” *J. Chromatogr. A*, vol. 1107, pp. 139–151, 2006.

[46] L. Nováková, L. Matysová, and P. Solich, “Advantages of application of UPLC in pharmaceutical analysis,” *Talanta*, vol. 68, pp. 908–918, 2006.

[47] A. Motoyama and J. R. Yates, “Multidimensional LC separations in shotgun proteomics,” *Anal. Chem.*, vol. 80, no. 19, pp. 7187–7193, 2008.

[48] G. J. Opiteck, S. M. Ramirez, J. W. Jorgenson, and M. A. Moseley, “Comprehensive two-dimensional high-performance liquid chromatography for the isolation of overexpressed proteins and proteome mapping.,” *Anal. Biochem.*, vol. 258, pp. 349–361, 1998.

[49] Y. Zhao, R. P. W. Kong, G. Li, M. P. Y. Lam, C. H. Law, S. M. Y. Lee, H. C. Lam, and I. K. Chu, “Fully automatable two-dimensional hydrophilic interaction liquid chromatography-reversed phase liquid chromatography with online tandem mass spectrometry for shotgun proteomics,” *J. Sep. Sci.*, vol. 35, pp. 1755–1763, 2012.

[50] R. J. C. Slebos, J. W. C. Brock, N. F. Winters, S. R. Stuart, M. A. Martinez, M. Li, M. C. Chambers, L. J. Zimmerman, A. J. Ham, D. L. Tabb, and D. C. Liebler, “Evaluating od strong cation exchange versus Icoelectric focusing of peptides for multidimensional liquid chromatography-yandem mass spectrometry,” *J. Proteome Res.*, vol. 7, pp. 5286–5294, 2008.

[51] D. A. Wolters, M. P. Washburn, and J. R. Yates, “An automated multidimensional protein identification technology for shotgun proteomics,” *Anal. Chem.*, vol. 73, no. 23, pp. 5683–5690, 2001.

[52] B. Manadas, J. A. English, K. J. Wynne, D. R. Cotter, and M. J. Dunn, “Comparative analysis of OFFGel, strong cation exchange with pH gradient, and RP at high pH for first-dimensional separation of peptides from a membrane-enriched protein fraction,” *Proteomics*, vol. 9, pp. 5194–5198, 2009.

[53] N. A. Patel, A. Crombie, S. E. Slade, K. Thalassinos, C. Hughes, J. B. Connolly, J. Langridge, J. C. Murrell, and J. H. Scrivens, “Comparison of one- and two-dimensional liquid chromatography approaches in the label-free quantitative analysis of Methylocella silvestris,” *J. Proteome Res.*, vol. 11, pp. 4755–4763, 2012.

[54] P. H. O’Farrell, “High Resolution of Proteins * Electrophoresis,” *J. Biol. Chem.*, vol. 250, no. 10, pp. 4007–4021, 1975.

[55] R. J. Braun, N. Kinkl, M. Beer, and M. Ueffing, “Two-dimensional electrophoresis of membrane proteins,” *Anal. Bioanal. Chem.*, vol. 389, pp. 1033–1045, 2007.

[56] A. E. Speers and C. C. Wu, “Proteomics of integral membrane proteins--theory and application.,” *Chem. Rev.*, vol. 107, pp. 3687–3714, 2007.

[57] S.-E. Ong, B. Blagoev, I. Kratchmarova, D. B. Kristensen, H. Steen, A. Pandey, and M. Mann, “Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.,” *Mol. Cell. Proteomics*, vol. 1, pp. 376–386, 2002.

[58] A. Konzer, A. Ruhs, H. Braun, B. Jungblut, T. Braun, and M. Krüger, “Stable isotope labeling in zebrafish allows in vivo monitoring of cardiac morphogenesis.,” *Mol. Cell. Proteomics*, vol. 12, pp. 1502–12, 2013.

[59] A. Sowell, K. E. Hersberger, T. C. Kaufman, and D. E. Clemmer, “Examining the Proteome of Drosophila Across Organism Lifespan research articles,” *J. Proteome Res.*, pp. 3637 – 3647, 2007.

[60] T. Geiger, J. R. Wisniewski, J. Cox, S. Zanivan, M. Kruger, Y. Ishihama, and M. Mann, “Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics.,” *Nat. Protoc.*, vol. 6, no. 2, pp. 147–157, 2011.

[61] T. Geiger, J. Cox, P. Ostasiewicz, J. R. Wisniewski, and M. Mann, “Super-SILAC mix for quantitative proteomics of human tumor tissue.,” *Nat. Methods*, vol. 7, no. 5, pp. 383–385, 2010.

[62] S. J. Deeb, R. C. J. D’Souza, J. Cox, M. Schmidt-Supprian, and M. Mann, “Super-SILAC Allows Classification of Diffuse Large B-cell Lymphoma Subtypes by Their Protein Expression Profiles,” *Mol. Cell. Proteomics*, vol. 11, pp. 77–89, 2012.

[63] L. Dayon, A. Hainard, V. Licker, N. Turck, K. Kuhn, D. F. Hochstrasser, P. R. Burkhard, and J. C. Sanchez, “Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags,” *Anal. Chem.*, vol. 80, no. 8, pp. 2921–2931, 2008.

[64] P. L. Ross, Y. N. Huang, J. N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlet-Jones, F. He, A. Jacobson, and D. J. Pappin, “Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.,” *Mol. Cell. Proteomics*, vol. 3, pp. 1154–1169, 2004.

[65] A. F. M. Altelaar, C. K. Frese, C. Preisinger, M. L. Hennrich, A. W. Schram, H. T. M. Timmers, A. J. R. Heck, and S. Mohammed, “Benchmarking stable isotope labeling based quantitative proteomics,” *J. Proteomics*, vol. 88, pp. 14–26, 2012.

[66] J. Villanueva, M. Carrascal, and J. Abian, “Isotope dilution mass spectrometry for absolute quantification in proteomics: Concepts and strategies,” *J. Proteomics*, vol. 96, pp. 184–199, 2014.

[67] W. I. Burkitt, C. Pritchard, C. Arsene, A. Henrion, D. Bunk, and G. O’Connor, “Toward Système International d’Unité-traceable protein quantification: From amino acids to proteins,” *Anal. Biochem.*, vol. 376, pp. 242–251, 2008.

[68] T. Huang, X. Zhang, D. Song, L. Wu, X. Li, Y. Yang, C. Quan, and H. Li, “The quantification of human chorionic gonadotropin by two isotope dilution mass spectrometry methods,” *Anal. Methods*, vol. 6, pp. 8690–8697, 2014.

[69] J. M. Pratt, D. M. Simpson, M. K. Doherty, J. Rivers, S. J. Gaskell, and R. J. Beynon, “Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes.,” *Nat. Protoc.*, vol. 1, no. 2, pp. 1029–1043, 2006.

[70] Z. Zhou, L. J. Licklider, S. P. Gygi, and R. Reed, “Comprehensive proteomic analysis of the human spliceosome.,” *Nature*, vol. 419, pp. 182–185, 2002.

[71] H. Liu, H. Liu, R. G. Sadygov, R. G. Sadygov, J. R. Yates, and J. R. Yates, “A model for random sampling and estimation of relative protein abundance in shotgun proteomics.,” *Anal. Chem.*, vol. 76, no. 14, pp. 4193–201, 2004.

[72] W. M. Old, K. Meyer-Arendt, L. Aveline-Wolf, K. G. Pierce, A. Mendoza, J. R. Sevinsky, K. A. Resing, and N. G. Ahn, “Comparison of label-free methods for quantifying human proteins by shotgun proteomics.,” *Mol. Cell. Proteomics*, vol. 4, pp. 1487–1502, 2005.

[73] Y. Ishihama, Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber, and M. Mann, “Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein.,” *Mol. Cell. Proteomics*, vol. 4, pp. 1265–1272, 2005.

[74] R. E. Higgs, M. D. Knierman, V. Gelfanova, J. P. Butler, and J. E. Hale, “Comprehensive label-free method for the relative quantification of proteins from biological samples,” *J. Proteome Res.*, vol. 4, pp. 1442–1450, 2005.

[75] J. M. Asara, H. R. Christofk, L. M. Freimark, and L. C. Cantley, “A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen,” *Proteomics*, vol. 8, pp. 994–999, 2008.

[76] G. Wang, W. W. Wu, W. Zeng, C. L. Chou, and R. F. Shen, “Label-free protein quantification using LC-coupled ion trap or FT mass spectrometry: Reproducibility, linearity, and application with complex proteomes,” *J. Proteome Res.*, vol. 5, pp. 1214–1223, 2006.

[77] J. Cox, M. Y. Hein, C. A. Luber, I. Paron, N. Nagaraj, and M. Mann, “Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ,” *Mol. Cell. Proteomics*, vol. 13, pp. 2513–2526, 2014.

[78] B. Schwanhäusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach, “Global quantification of mammalian gene expression control.,” *Nature*, vol. 473, pp. 337–342, 2011.

[79] J. R. Wiśniewski, P. Ostasiewicz, K. Duś, D. F. Zielińska, F. Gnad, and M. Mann, “Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma,” *Mol. Syst. Biol.*, vol. 8, no. 611, 2012.

[80] J. R. Wiśniewski, M. Y. Hein, J. Cox, and M. Mann, “A ‘ proteomic ruler ’ for protein copy number and concentration estimation without spike - in standards,” *Mol. Cell. Proteomics*, vol. 13, no. 12, pp. 3497–3506, 2014.

[81] J. C. Silva, M. V Gorenstein, G.-Z. Li, J. P. C. Vissers, and S. J. Geromanos, “Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition.,” *Mol. Cell. Proteomics*, vol. 5, pp. 144–156, 2006.

[82] M. Wilhelm, J. Schlegl, H. Hahne, A. Moghaddas Gholami, M. Lieberenz, M. M. Savitski, E. Ziegler, L. Butzmann, S. Gessulat, H. Marx, T. Mathieson, S. Lemeer, K. Schnatbaum, U. Reimer, H. Wenschuh, M. Mollenhauer, J. Slotta-Huspenina, J.-H. Boese, M. Bantscheff, A. Gerstmair, F. Faerber, and B. Kuster, “Mass-spectrometry-based draft of the human proteome.,” *Nature*, vol. 509, no. 7502, pp. 582–7, May 2014.

[83] J. C. Silva, R. Denny, C. A. Dorschel, M. Gorenstein, I. J. Kass, G.-Z. Li, T. McKenna, M. J. Nold, K. Richardson, P. Young, and S. Geromanos, “Quantitative proteomic analysis by accurate mass retention time pairs.,” *Anal. Chem.*, vol. 77, no. 7, pp. 2187–200, 2005.

[84] R. P. Dator, K. W. Gaston, and P. A. Limbach, “Multiple Enzymatic Digestions and Ion Mobility Separation Improve Quanti fi cation of Bacterial Ribosomal Proteins by Data Independent Acquisition Liquid Chromatography − Mass Spectrometry,” *Anal. Chem.*, vol. 86, no. 9, pp. 4264–4270, 2015.

[85] D. D. L. Tabb, L. Vega-Montoto, P. A. Rudnick, A. M. Variyath, A.-J. L. Ham, D. M. Bunk, L. E. Kilpatrick, D. D. Billheimer, R. K. Blackman, H. L. Cardasis, S. A. Carr, K. R. Clauser, J. D. Jaffe, K. A. Kowalski, T. A. Neubert, F. E. Regnier, B. Schilling, T. J. Tegeler, M. Wang, P. Wang, J. R. Whiteaker, L. J. Zimmerman, S. J. Fisher, B. W. Gibson, C. R. Kinsinger, M. Mesri, H. Rodriguez, S. E. Stein, P. Tempst, A. G. Paulovich, D. C. Liebler, and C. Spiegelman, “Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry.,” *J. Proteome Res.*, vol. 9, pp. 761–76, 2010.

[86] A. Bharti, P. C. Ma, and R. Salgia, “Biomarker discovery in lung cancer--promises and challenges of clinical proteomics.,” *Mass Spectrom. Rev.*, vol. 26, no. 3, pp. 451–466, 2014.

[87] N. Selevsek, C.-Y. Chang, L. C. Gillet, P. Navarro, O. M. Bernhardt, L. Reiter, R. Aebersold, L.-Y. Cheng, and O. Vitek, “Reproducible and consistent quantification of the Saccharomyces cerevisiae proteome by SWATH-MS,” *Mol. Cell. Proteomics*, vol. 14, pp. 739–749, 2015.

[88] B. C. Collins, L. C. Gillet, G. Rosenberger, H. L. Röst, A. Vichalkovski, M. Gstaiger, and R. Aebersold, “Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system.,” *Nat. Methods*, vol. 10, no. 12, pp. 1246–53, 2013.

[89] L. C. Gillet, P. Navarro, S. Tate, H. Rost, N. Selevsek, L. Reiter, R. Bonner, and R. Aebersold, “Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis,” *Mol. Cell. Proteomics*, vol. 11, no. 6, pp. O111.016717–O111.016717, 2012.

[90] S. J. Geromanos, J. P. C. Vissers, J. C. Silva, C. A. Dorschel, G.-Z. Li, M. V Gorenstein, R. H. Bateman, and J. I. Langridge, “The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS.,” *Proteomics*, vol. 9, no. 6, pp. 1683–95, 2009.

[91] C. S. Hoaglund-Hyzer, J. Li, and D. E. Clemmer, “Mobility labeling for parallel CID of ion mixtures,” *Anal. Chem.*, vol. 72, no. 13, pp. 2737–2740, 2000.

[92] C. Masselon, G. A. Anderson, B. R. Harkewicz, J. E. Bruce, L. Pasa-Tolic and R. D. Smith, "Accurate mass multiplexed tandem mass spectrometry for high-throughput polypeptude identification for mixtures, *Anal Chem,* vol. 11, pp. 1918-1924, 2000

[93] J. D. Venable, M. Q. Dong, J. Wohlschlegel, A. Dillin and J. R. Yates, "Automated appraoch for quantitative analysis of complex mixtures from tandem mass spectra, *Nat Methods,* vol. 1, pp. 39-45, 2004

[94] A. Panchaud, A. Scherl, S. A. Shaffer, P. D. von Haller, H. D. Kulasekara, S. I. Miller and D. R. Goodlett, "Precursor acquisition indepednet from ion count: how to dive deeper into the proteomics ocean, *Anal. Chem.,* vol. 81, pp. 6481-6488, 2009

[95] S. D. Pringle, K. Giles, J. L. Wildgoose, J. P. Williams, S. E. Slade, K. Thalassinos, R. H. Bateman, M. T. Bowers, and J. H. Scrivens, “An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument,” *Int. J. Mass Spectrom.*, vol. 261, no. 1, pp. 1–12, 2007.

[96] E. Rodriguez-Suarez, C. Hughes, L. Gethings, K. Giles, J. Wildgoose, M. Stapels, K. E. Fadgen, S. J. Geromanos, P. C. J. Vissers, F. Elortza, and J. I. Langridge, “An ion mobility assisted data independent LC-MS strategy for the analysis of complex biological samples,” *Curr. Anal. Chem.*, vol. 9, no. 2, pp. 199–211, 2013.

[97] K. Thalassinos, S. E. Slade, K. R. Jennings, J. H. Scrivens, K. Giles, J. Wildgoose, J. Hoyes, R. H. Bateman, and M. T. Bowers, “Ion mobility mass spectrometry of proteins in a modified commercial mass spectrometer,” *Int. J. Mass Spectrom.*, vol. 236, no. 1–3, pp. 55–63, 2004.

[98] B. McCullough, J. Kalapothakis, H. Eastwood, P. Kemper, D. MacMillan, K. Taylor, J. Dorin, and P. Barran, “Development of an ion mobility quadrupole time of flight mass spectrometer,” *Anal. Chem.*, vol. 80, no. 16, pp. 6336–6344, 2008.

[99] C. S. Hoaglund, S. J. Valentine, and D. E. Clemmer, “An Ion Trap Interface for ESI−Ion Mobility Experiments,” *Anal. Chem.*, vol. 69, no. 20, pp. 4156–4161, 1997.

[100] B. T. Ruotolo, J. L. P. Benesch, A. M. Sandercock, S.-J. Hyung, and C. V Robinson, “Ion mobility-mass spectrometry analysis of large protein complexes.,” *Nat. Protoc.*, vol. 3, no. 7, pp. 1139–52, 2008.

[101] P. V. Shliaha, N. J. Bond, L. Gatto, and K. S. Lilley, “The Effects of Travelling Wave Ion Mobility Separation on Data Independent Acquisition in Proteomics Studies.,” *J. Proteome Res.*, 2013.

[102] U. Distler, J. Kuharev, P. Navarro, Y. Levin, H. Schild, and S. Tenzer, “Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics.,” *Nat. Methods*, vol. 11, no. 2, pp. 167–70, 2014.

[103] H. Xu and M. A. Freitas, “A mass accuracy sensitive probability based scoring algorithm for database searching of tandem mass spectrometry datA.,” *BMC Bioinformatics*, vol. 8, p. 133, 2007.

[104] A. Keller, A. I. Nesvizhskii, E. Kolker, and R. Aebersold, “Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search,” *Anal. Chem.*, vol. 74, no. 20, pp. 5383–5392, 2002.

[105] N. Zhang, R. Aebersold, and B. Schwikowski, “ProbID: A probabilistic algorithm to identify peptides through sequence database searching using tandem mass spectral data,” *Proteomics*, vol. 2, no. 10, pp. 1406–1412, 2002.

[106] J. E. Elias and S. P. Gygi, “Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry.,” *Nat. Methods*, vol. 4, no. 3, pp. 207–214, 2007.

[107] E. Bonzon-kulichenko, F. Garcia-marques, and M. Trevisan-herraz, “Revisiting Peptide Identi fi cation by High-Accuracy Mass Spectrometry: Problems Associated with the Use of Narrow Mass Precursor Windows”, *J. Proteome. Res,* vol. 14, no. 2, pp. 700-710, 2015.

[108] G.-Z. Li, J. P. C. Vissers, J. C. Silva, D. Golick, M. V Gorenstein, and S. J. Geromanos, “Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures.,” *Proteomics*, vol. 9, no. 6, pp. 1696–719, 2009.

[109] D. N. Perkins, D. J. C. Pappin, D. M. Creasy, and J. S. Cottrell, “Probability-based protein identification by searching sequence databases using mass spectrometry data,” *Electrophoresis*, vol. 20, no. 18, pp. 3551–3567, 1999.

[110] J. Cox, N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen, and M. Mann, “Andromeda: A peptide search engine integrated into the MaxQuant environment,” *J. Proteome Res.*, vol. 10, no. 4, pp. 1794–1805, 2011.

[111] R. Craig and R. C. Beavis, “TANDEM: Matching proteins with tandem mass spectra,” *Bioinformatics*, vol. 20, no. 9, pp. 1466–1467, 2004.

[112] L. Y. Geer, S. P. Markey, J. A. Kowalak, L. Wagner, M. Xu, D. M. Maynard, X. Yang, W. Shi and S.H. Bryant “Open mass spectrometry search algorithm” *Journal of proteome research*, *3*(5), 958-964,2004.

[113] J. K. Eng, A L. McCormack, and J. R. Yates, “An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database.,” *J. Am. Soc. Mass Spectrom.*, vol. 5, no. 11, pp. 976–89, 1994.

[114] J. Cox and M. Mann, “MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.,” *Nat. Biotechnol.*, vol. 26, no. 12, pp. 1367–72, 2008.

[115] T. Kind, G. Wohlgemuth, D. Y. Lee, Y. Lu, M. Palazoglu, S. Shahbaz, and O. Fiehn, “FiehnLib: Mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry,” *Anal. Chem.*, vol. 81, no. 24, pp. 10038–10048, 2009.

[116] G. Smith, Colin A; Maille, Grace O'; Want, Elizabeth J; Qin, Chuan; Trauger, Sunia A; Brandon, Theodore R; Custodio, Darlene E; Abagyan, Ruben; Siuzdak, "METLIN: A Metabolite Mass Spectral DatabaseTitle," *Ther. Drug Monit.*, vol. 27, no. 6, pp. 747–751, 2005.

[117] H. Lam, E. W. Deutsch, J. S. Eddes, J. K. Eng, S. E. Stein, and R. Aebersold, "Building consensus spectral libraries for peptide identification in proteomics.," *Nat. Methods*, vol. 5, no. 10, pp. 873–875, 2008.

[118] K. Thalassinos, J. P. C. Vissers, S. Tenzer, Y. Levin, J. W. Thompson, D. Daniel, D. Mann, M. R. DeLong, M. A. Moseley, A. H. America, A. K. Ottens, G. S. Cavey, G. Efstathiou, J. H. Scrivens, J. I. Langridge, and S. J. Geromanos, "Design and application of a data-independent precursor and product ion repository," *J. Am. Soc. Mass Spectrom.*, vol. 23, no. 10, pp. 1808–1820, 2012.

[119] D. N. Hebert and M. Molinari, "In and out of the ER: protein folding, quality control, degradation, and related human diseases.," *Physiol. Rev.*, vol. 87, no. 4, pp. 1377–1408, 2007.

[120] B. Bukau, J. Weissman, and A. Horwich, "Molecular Chaperones and Protein Quality Control," *Cell*, vol. 125, no. 3, pp. 443–451, 2006.

[121] J. King, C. Haase-Pettingell, and D. Gossard, "Protein folding and misfolding," *Am. Sci.*, vol. 90, no. 5, pp. 445–453, 2002.

[122] P. Westermark, K. Sletten, B. Johansson, and G. G. Cornwell, "Fibril in senile systemic amyloidosis is derived from normal transthyretin.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 87, no. 7, pp. 2843–2845, 1990.

[123] R. E. Tanzi and L. Bertram, "Twenty years of the Alzheimer's disease amyloid hypothesis: A genetic perspective," *Cell*, vol. 120, no. 4, pp. 545–555, 2005.

[124] E. H. Koo, P. T. Lansbury, and J. W. Kelly, "Amyloid diseases: abnormal protein aggregation in neurodegeneration.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 96, no. 18, pp. 9989–9990, 1999.

[125] R. Van Der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Old, R. V Pappu, P. Tompa, V. N. Uversky, P. E. Wright, and M. M. Babu, "Classification of Intrinsically Disordered Regions and Proteins," *Chem. Rev.*, vol. 114, no. 13, pp. 6589–6631, 2014.

[126] M. H. Polymeropoulos, C. Lavedan, E. Leroy, S. E. Ide, A. Dehejia, A. Dutra, B. Pike, H. Root, J. Rubenstein, R. Boyer, E. S. Stenroos, S. Chandrasekharappa, A. Athanassiadou, W. G. Johnson, A. M. Lazzarini, R. C. Duvoisin, G. Di Iorio, L. I. Golbe, R. L. Nussbaum, E. Reports, C. Lavedant, E. Leroyt, T. Papapetropoulos, G. Di, and L. Golbe, "Mutation in the alpha-Synuclein Gene Identified in Families with Parkinson ' s Disease," *Science*, vol. 276, no. June, pp. 2045–2048, 2012.

[127] K. Fargo, "Alzheimer's Association Report: 2014 Alzheimers disease facts and figures," *Alzheimer's Dement.*, vol. 10, no. 2, pp. e47–e92, 2014.

[128] J. Lund, P. Tedesco, K. Duke, J. Wang, S. K. Kim, and T. E. Johnson, "Transcriptional profile of aging in C. elegans," *Curr. Biol.*, vol. 12, no. 18, pp. 1566–1573, 2002.

Data Independent Quantitative Proteomic Analysis of Age-Related Vulnerability to Aβ42 Toxicity in *D. melanogaster*

2.0 Abstract

Alzheimer's disease (AD) accounts for 60-80% of all senile dementia cases and is one of the fastest growing chronic diseases in developed countries. Although the aetiology of the disease has been linked to increased levels of Aβ peptide concentration within the brain and the subsequent formation of toxic protein aggregates, the biggest risk factor for developing the disease is increasing age. Previously the use of an inducible *D. melanogaster,* for the first time, successfully demonstrated age-related vulnerability to the neuronal expression of an aggregating variant of Aβ42 peptide. Using a data independent, mass spectrometric approach, quantitative proteomic analysis of *D. melanogaster* brain tissue at two different ages, with and without Aβ42 induction, was undertaken to identify proteins implicated in age-related vulnerability to peptide aggregation. Using our experimental design, differentially expressed proteins were successfully linked to Aβ toxicity, increasing age and a combination of the two variables. Results indicate a mixed response of chaperones to both Aβ and age which had all previously been identified as modulators of peptide toxicity and longevity. Our data provides further evidence towards the possible implication of altered chaperone activity in age-related susceptibility to Alzheimer's disease. Interestingly protein-protein interaction analysis identified enrichment in proteins involved in small GTPase mediated signal transduction to be regulated across conditions. Here, complex and varied expression levels were measured indicating mixed responses to age and Aβ42 aggregation. Our unbiased approach has provided a novel insight into cellular responses to both ageing and Aβ peptide toxicity. These data provide avenues for follow up functional studies, potentially with the same model organism, to better understand the link between ageing and Alzheimer's disease.

2. 1 Introduction

Alzheimer's disease is the most common form of senile dementia, accounting for between 60-80% of all cases and represents one of the fastest growing chronic and fatal diseases in developed countries [1]. The cause and progression of the disease is not well understood but the widespread presence of extracellular amyloid plaques in the brain tissue of AD patients, supports that the aetiology of the disease is an increase of amyloid beta (Aß) peptide concentration within the brain [2], [3]. These Aβ peptides vary in length (between 38-43 amino acids long) and are therefore referred to in reference to their length, i.e. Aβ42. The peptide species are known to be derived from the amyloid precursor protein (APP) following it sequential cleavage by the proteolytic enzymes α-, β- and γ-secretase, **Figure 2.1**, however their normal biological function is unknown. Indeed neurodegenerative diseases are almost exclusively linked with the formation of toxic protein aggregates, often within neurons. Increased Aß concentration is linked by a large body of evidence, to Aβ oligomerisation, leading to aggregation and the formation of extended aggregate structures, amyloid plaques, in the extracellular regions that surround neurons.[4]

The 'amyloid cascade' hypothesis is further supported by the mutations that have been identified as a cause of the familial form of Alzheimer's disease (FAD). Mutations to the amyloid precursor protein, from which Aß peptides are enzymatically derived, and to the presenilins PS1 and PS2, which are both involved in the enzymatic cleavage process, are directly implicated in the familial form of AD [5], [6]. Such mutations lead to an increase in Aß production, the propensity for Aß42 to aggregate, or levels of Aß42 relative to Aß40. A large body of research into AD has focused primarily on the role of amyloid aggregates, using genetic models based upon FAD. These models range from *C. elegans* through to *M. musculus* and display AD-related phenotypes including shortened lifespan, cognitive impairment and/or neuronal dysfunction and the accumulation of Aß in an age-dependent manner [7]–[9]. FAD only accounts for <1% of all AD cases and whilst genetic wide association studies (GWAS) have successfully identified mutations associated with the development of sporadic AD, age is still the biggest risk factor [10]–[12]. Age as a risk factor is common among many neurodegenerative diseases with increasing age also seen to increase the propensity for developing Parkinson's disease [13]. The link between age and neurodegenerative diseases, however, is not well understood mechanistically. Elucidating the mechanisms that link AD and age so closely should aid in the

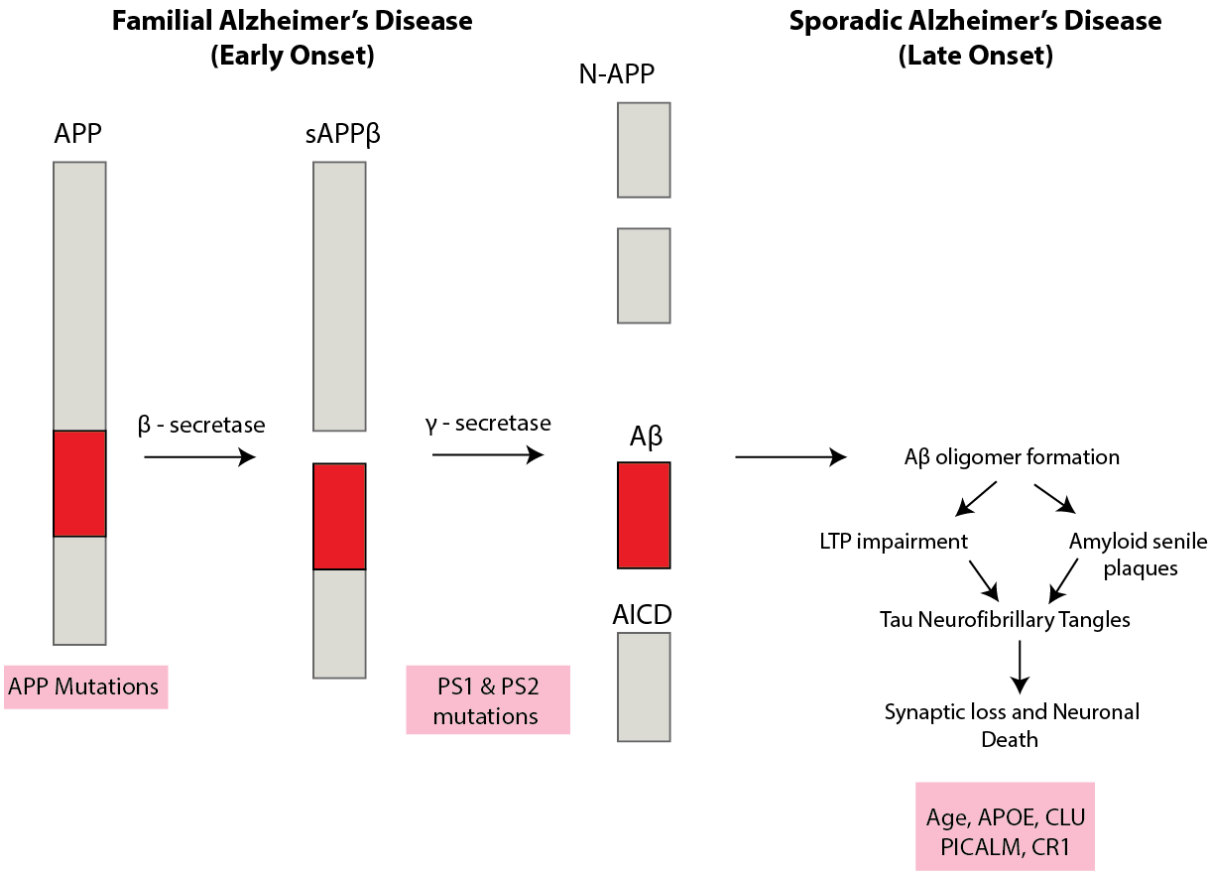


Figure 2.1 The processing of the amyloid precursor protein and it's relation to Alzheimer's Disease. APP is cleaved by β-secretase and subsequently by γ-secretase to form the Aβ peptide, of which, an increase in concentration is thought to start the amyloidogenic pathway. Mutations to APP and the transmembrane proteins presenilin 1 & 2 (PS1 & 2) are known to cause familial Alzheimer's disease through an increase in Aβ peptide concentration. In sporadic AD, however, these mutations are not present and the disease occurs at a much later age. The disease is also thought to be as a result of increased Aβ peptide concentration, although the cause for this is unknown. A number of genes are known to contribute to increased susceptibility to sporadic AD (pink box), however the biggest risk factor is increased age.

development of improved treatments and therapies that up until now have been unsuccessful [14]. Such research is particularly pertinent as a result of rapidly ageing populations in both developed and developing countries, which is driving the increased prevalence of AD. Currently it is estimated that >40 million people are living with dementia and this figure is predicted to rise to 135.5 million, with a large proportion of these cases being from poorer developing countries [12].

Recently, with the use of an inducible adult-onset *Drosophila* AD model, **Figure 2.2,** an age related vulnerability to the expression of the FAD -associated Arctic Aß42 (Arc-Aß42) isoform, specifically in neuronal cells was reported, **Figure 2.3** [15]. Normalised expression of Arc-Aß42 was induced in flies of three different ages for equivalent time periods. After controlling for the duration of exposure to the toxic peptide and for normal ageing, it was shown through both negative geotaxis assays (climbing assay) and lifespan measurements, that older flies were significantly more susceptible to peptide toxicity. The underlying mechanisms behind the observed vulnerability were beyond the scope of the initial paper and have subsequently been pursued using an unbiased mass spectrometry-based proteomics approach.

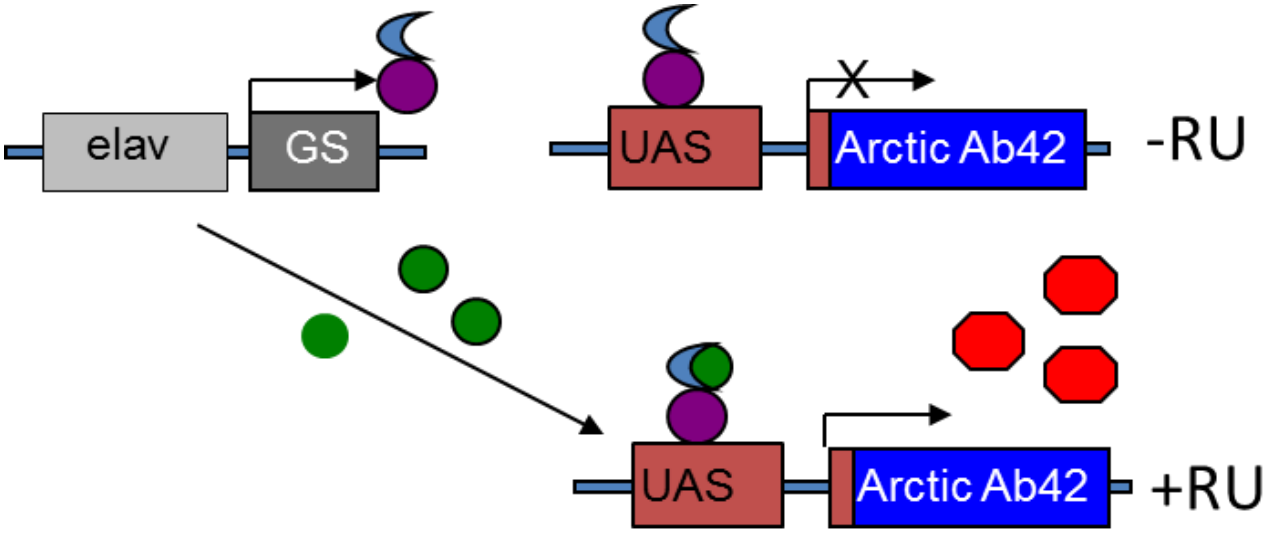


Figure 2.2 Representation of the GeneSwitch-UAS expression system used in adult onset Alzheimer *Drosophila* models.[7] Transgenic flies contain the transcriptional activator GeneSwitch which is under control of the nervous system specific elav promoter, termed elavGS (in grey). An Arctic Aβ42 transgene is fused to a GAL-4 binding upstream activation sequence , UAS-Aβ (red and blue). During treatment with the hormonal rug mifepristone RU486 (+RU, Green), the GeneSwitch protein become transcriptionally active, binds to UAS and induces the expression of Aβ42 within the nervous system. Without RU486 treatment, the gene switch protein is still expressed but is transcriptionally silent.

Mass spectrometry (MS)-based proteomics, in particular the identification of differentially expressed proteins, through both relative and absolute quantification has been an important technique to help elucidate the constituent pathways that may contribute to AD progression. The majority of this

research has focused on the identification of enriched or depleted proteins in the post-mortem brains of AD patients, leading to a better understanding of the late-stages of the disease. Such proteomic studies have helped confirm the roles of inflammation and oxidative stress, amongst others, in the mechanism of neuronal and synaptic dysfunction. Information attainable from post-mortem samples is however, limited to the final stages of the disease, a point at which, it is thought, possible therapeutic routes will be small in number. The use of animal AD models alongside quantitative proteomics facilitates the collection of data that can better explain the earlier stages of the disease progression, where the identification of druggable targets has a better chance to prove effective. One such unbiased proteomics approach which utilized a triple transgenic AD mouse model was able to directly implicate mitochondrial dysfunction with Aβ and tau expression. Follow up experiments then effectively validated the synergistic effects of Aβ and tau protein in mitochondrial death, which is a well characterized AD pathogenic pathway [16].

The mainstay of quantitative proteomics has been the use of label-based methods, notably stable isotope labelling by amino acids in cell culture (SILAC) and isobaric labelling methods such as tandem mass tag (TMT) or isobaric tag for relative and absolute quantification (iTRAQ) [17]–[19]. Although these methods offer high precision, in part due to the minimization of errors during sample preparation and mass spectrometric analysis, they also have their limitations. SILAC is now routine for *in vitro* screens however its use in model organisms is not always straightforward. Until recently near complete incorporation of heavy amino acids in *Drosophila* was not possible. Since its emergence as a viable technique in *Drosophila*, little research has been published with the use of such a method [20]. Recent developments that have provided robust liquid chromatography mass spectrometry platforms, alongside increased linear dynamic range and high duty cycle have meant that label-free quantitative proteomics has seen a surge in popularity, as it has become a viable alternative for quantitative studies [21], [22]. Label free proteomics now can routinely produce technical reproducibility <10% coefficient of variance and with the addition of improved normalization techniques has been reported to be as low as 1% [23], [24]. The benefits of label free-based methods include low cost, the ability to directly compare an unlimited number of samples and improved proteome coverage in comparison to label-based method where the multiplexing of analysis increases sample complexity.

As the popularity of label-free has increased, a multitude of approaches both in terms of acquisition methods and data analysis have been developed, aimed at improved precision and accuracy of quantitation [25], [26]. Data independent acquisition (DIA) strategies have emerged as an alternate approach to the classical mass spectrometry-based proteomic approach of data dependent acquisition (DDA), where precursors are selected for fragmentation based upon intensity (see Chapter 1.16) [27].

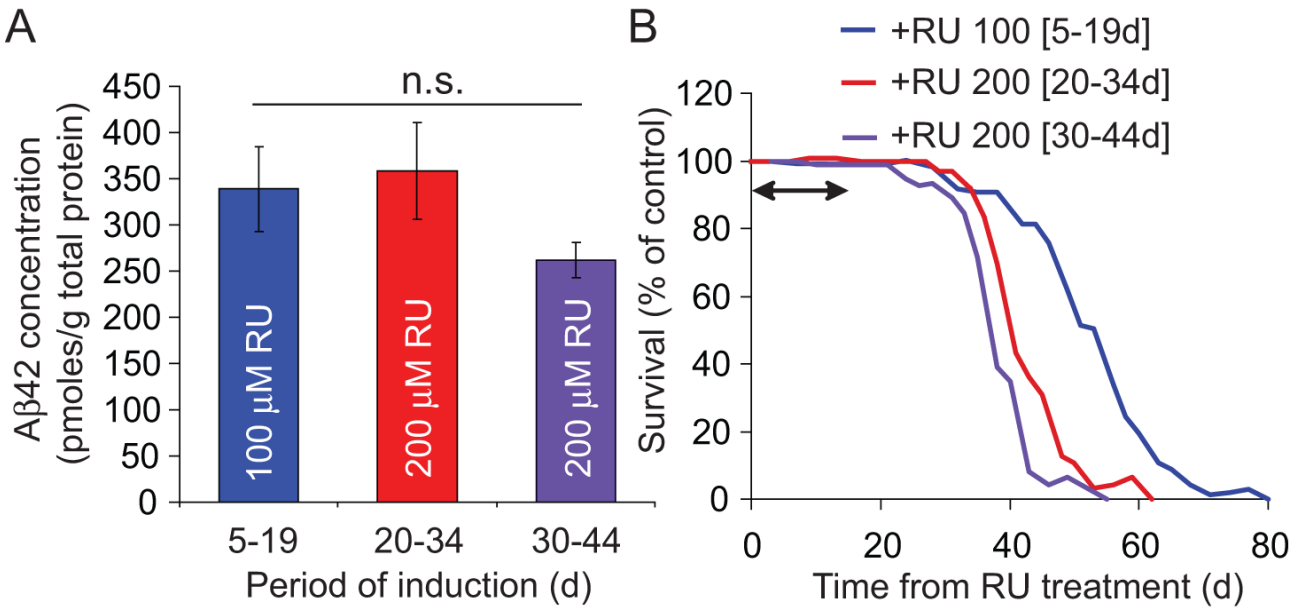


Figure 2.3 Age-related vulnerability to Aβ42 in *D. melanogaster.* **A)** Standardised Aβ42 expression, induced in flies of different ages. RU486 concentration was varied to account for differences in feeding and protein translation. **B)** Survival curves with Aβ42 induction carried out at three different ages. Induction in flies 5-19 days old demonstrates a statistically significant increase in lifespan compared to flies with Aβ42 induction at 20-34 days or 30-44 days.

The suitability of MS^E for quantitative proteomic studies has been well demonstrated by a number of applications. A well designed evaluation by Levin et. al, demonstrated a linear dynamic range of 2.5-3 orders in samples of varying complexity, with R^2 values >0.99 consistently reported [28]. The technique has been used with success to study a number of different biological systems including AD. M.D. Hoos and colleagues were able to identify a range of differentially expressed proteins extracted from the brains of 42 week old CVN-AD mice; a transgenic mouse line that presents age-related accumulation of Aβ under low nitric oxide conditions [29].

In order to probe the underlying biological system perturbations that may help explain the age-related vulnerability to Aβ toxicity previously reported, we employed a LC-MS^E platform with prior gel-based fractionation, to probe in an unbiased manner the underlying pathways that might contribute towards the age-related vulnerability to Aβ toxicity observed in *Drosophila*. Our experimental design facilities the identification of proteins that are differentially expressed upon Aß42 expression, with age or display both an Aβ and age-related response. Proteins identified as differentially expressed from binary comparisons were found to carry out varied cellular functions, as assigned by gene ontology terms. Protein-protein interaction network analysis highlighted an enrichment of differentially expressed small GTPase proteins in older flies. Expression profiling found these proteins to display both age and Aβ expression responses, however the direction of expression change across this set of proteins was varied.

2.2 Materials and Methods

2.2.1 Fly stocks and Husbandry

Fly stocks were maintained at 25°C on a 12:12-h light and dark cycle. Humidity was kept constant and flies were fed on a sugar-yeast (SY) medium which consisted of 15 g l^{-1} agar, 50 g l^{-1} sugar, 100 g l^{-1} autolysed yeast, 100 g l^{-1} nipagin and 3ml l^{-1} propionic acid. Where neuronal expression of arctic Aβ42 is indicated, food was prepared as stated and spiked with mifepristone (RU486) at a final concentration of 200 µM.

The use of the elav GeneSwitch (elavGS)-UAS system [GAL4-dependant upstream activator sequence] allowed the adult-onset neuronal-specific expression of Arctic Aβ42 peptide upon treatment with the progesterone steroid, mifepristone (RU486, Sigma).[30] UAS and elavGS-lines had been backcrossed into the w^{1118} genetic background six times prior to male flies expressing UAS-constructs being crossed to female flies expressing elavGS to ensure a homogenous wild-type background genome.

2.2.2 Aβ42 Expression and Dissections

Fly crosses were staggered so that Aβ42 expression was induced in both 5 day and 20 day old *Drosophila* simultaneously. Expression was induced for 7 days for both conditions. *Drosophila* brain tissue was dissected 14 days post initial induction. For each biological repeat dissections were conducted at the same time each day, to limit experimental variability caused by circadian rhythm effects. A total of fifteen flies were dissected per sample. Flies were anaesthetized on CO_2 and dissections were carried out on ice-cold phosphate buffer saline solution and transferred immediately to RNAlater (Life technologies, USA) to prevent enzymatic protein degradation. Crosses, expression, ageing-spans and dissections were carried out as a collaborative effort with Dr Fiona Kerr, Healthy Ageing Department, UCL.

2.2.3 SDS-PAGE Separation and Tryptic Digestion.

Samples were prepared for SDS-polyacrylamide gel electrophoresis (PAGE) by homogenisation in 30 µl of x2 Laemmli sample buffer (Life technologies,USA) containing β-mercaptoethanol (Sigma,USA). Samples were then boiled for 5 minutes just prior to SDS-PAGE. Proteins were separated on pre-cast 10 % SDS-PAGE gels (Bio-Rad, USA and visualised by staining with Coomassie brilliant blue R-250 (Pierce, USA). Samples from the four conditions from the same biological repeat were batched on the same gel in order to reduce variability. Each of the 8 lanes were excised from the gel and cut into ten bands. The size of each band was decided based upon the amount of protein present, as indicated by the extent of staining. The size of each band was kept consistent across all samples prepared.

Bands were cut into small sections for efficient digestion, destained and dehydrated with acetonitrile. Disulphide bonds were reduced by covering gel pieces with 100 µl of 10 mM DTT (Sigma Aldrich, St.Louis, MO) with heating at 56°C for 30 minutes. Free thiols were alkylated with 100 µl of 55 mM IAA (Sigma Aldrich, St.Louis, MO). The reaction was left for 30 minutes in the dark. Gel pieces were washed with 50 mM ammonium bicarbonate (Sigma Aldrich, St.Louis, MO) and dehydrated with acetonitrile. This step was repeated twice. Gel pieces were rehydrated with 100 µl of a 10 ng ul^{-1} trypsin solution in 50 mM ammonium bicarbonate. Samples were left overnight at 37°C. Post digestion peptides were extracted by addition of 100 µl of extraction solution (97% water, 2% acetonitrile, 1% formic acid). The supernatant was collected and 70 µl of acetonitrile added to the gel pieces to extract remaining hydrophobic peptides. The supernatant was added to the previous solution and dried down by vacuum centrifugation. Samples were re-solubilised in 10 µl of 0.1 % formic acid. Samples were frozen at -20 °C prior to analysis.

2.2.4 Quantification of Total Aβ42

Total Aβ42 was extracted based upon a published method.[31] Fly heads were homogenised in 50 µl guanidine hydrochloride extraction buffer (5 M GnHCl, 50 mM Hepes pH 7.3, protease inhibitor cocktail (Sigma) and 5 mM EDTA). Homogenate was centrifuged at 21,000 g for 5 minutes at 4°C before extracting the supernatent for total Aβ42 measurment. Aβ42 concentration was measured using the High Sensitivity Human Amyloid Aβ42 ELISA kit (Millipore, Bilerica, MA, USA). Samples

were diluted 1:100 in sample dilution buffer and the immunoassay was performed according to the manufacturer's instructions. A Bradford protein assay (Bio-Rad laboratories Ltd, UK) was used to quantify total protein levels from each extraction to allow the calcualtion of Aβ42 concentration as a ratio to total protein (pmoles/g).

2.2.5 Reverse Phase Liquid-chromatography Mass Spectrometry Analysis

For each sample 1 µl was loaded using a split-less nano-Ultra Performance Liquid Chromatography (10 kpsi nanoAcquity Waters Corporation, Milford, MA, USA). Buffers used were (A) water + 0.1% formic acid (J.T. Baker, PA, USA) and (B) acetonitrile + 0.1% formic acid (J.T. Baker, PA, USA). Samples were desalted using a reverse-phase SYMMETRY C18 trap column (180 µm internal diameter, 20 mm length, 5 µm particle size, Waters Corporation) at a flow rate of 5 µl/min for 3 minutes. Peptides were separated by a linear gradient (0.3 µl/min, 40 °C; 97-60% Buffer A over 90 minutes) using a BEH130 C18 nano-column (75 µm internal diameter, 150 mm length, 1.7 µm particle size, Waters Corporation).

The nanoLC was coupled online through a custom built nanoflow sprayer to a quadrupole time-of-flight (QToF) hybrid mass spectrometer (HDMS Synapt; Waters UK) tuned to a mass resolution of 10 000 (full width half height). The ToF analyser was externally calibrated from *m/z* 175.11 to 1285.54 using the fragment ions of a 500 fmol/µl solution of [Glu^1]-fibrinopeptide B (GFP, Sigma Aldrich, St.Louis, MO). Data were lockmass-corrected post acquisition using the monoisotopic mass of the doubly charged precursor of GFP (785.8426 *m/z*), delivered at 500 fmol/µl to the mass spectrometer via a NanoLockSpray interface. The reference sprayer was sampled every 60s. Accurate mass measurements were made using a data independent mode (LC-MS^E) of acquisition.[32], [33] Briefly, energy in the collision cell was alternated from low energy (4 eV) to high energy (energy ramp from 15-35 eV) whilst continuously aquiring MS data. Measurements were made over a *m/z* range of 100-1990 Da with a scan time of 0.6 s and an interscan delay of 0.05s. One cycle of MS and MS^E data were acquired every 1.3 s. A radio frequency was applied to the quadrupole mass analyser to facilate the efficient transmission of ions with an *m/z* range of 300-2000. This ensured that ions below 300 *m/z* and observed in the MS^E spectrum were known to be derived from collision cell fragmentation.

Biological replicates were analysed in technical duplicate. Sample analysis was batched so that conditions from the same biological repeat were analysed during the same analytical run.

2.2.6 Database Processing

Databases were searched using PLGS v2.5.2 (Waters Corporation). The computational methods used to process the data are explained in detail by S.J. Geromanos [27]. Briefly raw data is lockmass-corrected, smoothed, background subtracted and deisotoped. This provides a single centriod accurate mass for the monoisotopic of each peptide and respective fragment ions. Peptide are associated with fragment ions through time aligment. Data were searched against an Uniprot complete protein database for *Drosophila melanogaster*. Carbamidomethyl-C was specified as a fixed modification. Oxidation (M), phosphorylation (S,T,Y), acetylation (K, N-terminus) and deamidation (N and Q) were specified as variable modifications. A maximum of two missed cleavages of the protease were allowed for semi-tryptic peptide identification. For peptide identification two corresponding fragment ions were set as a minimum criterion whereas for protein identification a minimum of one corresponding peptide identification and two fragment ions were required. Protein level false discovery rate was maintained at < 5% estimated based upon the number of proteins identified from a randomised database. The false discovery rate was minimised further by applying the additional criteria that for a successful identification a protein must be identified in a minimum of two biological repeats and three technical repeats. This criterion is based on the assumption that false positives, from chemical noise, are random in nature and should not replicate across injections. Database search results were outputted in .csv format for further analysis. These data are provides in the appendix file.

2.2.7 Protein Quantification

Relative protein quantification was achieved using the Hi3 technique as detailed elsewhere [34]. Briefly, it has been shown that the combined intensity for the three most intense peptides for each protein in a proteomics run directly correlates (R^2=0.99) to the protein amount loaded onto to column [35]. This observation allows the direct comparison of conditions without the need for isotopic labelling

or isobaric tagging techniques. To account for run to run instrumental variation, a normalisation technique was applied across runs to be compared.

2.2.8 Peptide Intensity Normalisation

Within condition normalisation was achieved by an in-house R script based upon locally weighted scatterplot smoothing (LOESS) and was conducted on data at the peptide level [36], [37]. LOESS-based normalisation is founded upon the assumption that when comparing conditions, even in the presence of differential expression, the majority of protein abundances stay constant. The result of this is that if samples are compared with an MA plot (where M is protein abundance ratio, between samples, and A is average protein intensity, both plotted on a $\log_2$ scale) the fit of a locally weighted regression line across the intensity range should perfectly dissect 0 on the y-axis. In reality this does not always hold true. The LOESS approach normalises data by applying correction values to raw peptide intensities such that the locally weighted regression line better fits the desired y=0 equation post correction. The script used to perform the normalisation is provided in **Appendix A.**

2.2.9 Effect Size Calculation

Effect size was calculated based upon the method used previously by M.D Hoos *et. al* in their label-free quantitative proteomic study of an Alzheimer's mouse model [29]. Briefly, a moving average of data variance across protein intensity distribution was calculated. Data variance was calculated as a coefficient of variation (%) for all proteins that had three or more quantitative values in a condition. Data from all conditions were included in the calculation and CVs from proteins identified in more than one condition were treated independently. A 100 point moving average was calculated. Using these data, effect size (as Cohen's d) was calculated for each quantified protein. Proteins taken forward for statistical analysis were required to have a calculated Cohen's d >1.0. This represents a difference in mean which is larger than one standard deviation of the calculated data variance.

2.2.10 Statistical Analysis

Data deemed significantly different, based upon effect size, were taken forward for statistical analysis. An unpaired Student's t-test, assuming equal variance, was carried out on log_2 transformed Hi3 protein intensities. Conditions were subject to pairwise comparison. Significance levels were chosen based upon the false discovery rate due to multiple hypothesis testing. P values deemed significant were chosen to ensure statistically less than one protein per comparison should be a false positive. An additional definition of differential expression was included which took advantage of missing protein identifications. During pairwise comparison if a protein was successfully quantified in one condition, based on previously mentioned criteria, but was not quantified in any of the replicates from the second condition, it was deemed to be differentially expressed. These data are provided in the appendix file.

2.2.11 Western Blotting

Drosophila heads were pooled and homogenised in 2x Laemmli sample buffer (Life technologies,USA) sample buffer containing β-mercaptoethanol (Sigma Aldrich, St.Louis, MO) and heated at 95°C for 5 minutes. Proteins were separated on SDS-PAGE gels and blotted onto nitrocellulose. Membranes were incubated in a blocking solution containing 5% milk protein in Tris-Buffered Saline-Tween20 (TBST) solution for 1 hour at room temperature. Primary antibodies were diluted in TBST according to previous literature or manufacturer's instructions (ranging between 1:500 & 1:2000) and blotted nitrocellulose were incubated in primary antibody overnight at 4°C. Following incubation with the appropriate HRP-conjugated secondary antibodies (Abcam, Cambridge, UK) for one hour at room temperature blots were developed using an ImageQuant LAS 4000 (GE Healthcare Life Sciences, Little Chalfont, UK). Quantitation was performed using ImageQuant software.

Primary antibodies used were Rab-3 (Dr Aaron DiAntonio), Ran (Cell Signaling Technology, Danvers, MA, USA), AP50 (BD Transduction Laboratories, New Jersey, USA), EndoA (Dr Patrik Verstreken)

2.3 Results

2.3.1 Comparative Analysis of the *Drosophila* Head and Brain Proteome

Few previous studies looking at global levels of protein or RNA upon neurodegeneration in *Drosophila*, have specifically analysed brain tissue. Commonly fly head homogenate is sampled due to the ease of obtaining large amounts of sample with a relatively quick protocol. This approach is based on the assumption that analysis of the head is representative of the brain. The *Drosophila* brain, however, only represents a proportion of the total head mass, with the eyes and cuticle contributing a large percentage of protein mass. We hypothesised that using a mass spectrometry workflow that is unable to identify the complete head proteome, to analyse head samples, might dramatically reduce the number of identifiable brain specific proteins. We carried out a direct comparison of between an extract from 10 *Drosophila* heads to an extract from 10 *Drosophila* brains, **Figure 2.4A**. Using our 1D-gel-MS^E workflow we identified similar numbers of proteins from the head and brain sample following triplicate analytical replicates. Proteins were classed as successfully identified if they were present in 2 out of 3 replicates. Based on the total number of proteins identified, 1221, around 1/3 were attributable as specifically expressed in the head whilst a similar number were only identified from the brain sample. Due to the stochastic nature of MS analysis the replicate analysis of a sample can identify different sets of proteins. In order to validate our results we examined whether the use of three technical replicates gave results that were consistent enough to draw comparisons between the two different samples, **Figure 2.4B.** A level of reproducibility was observed between technical replicates which provided confidence that results seen in **Figure 2.4A,** were truly representative of the samples analysed and were not a result of technical variations. Based on these data it was decided that the analysis of brain tissue would provide results more representative of age-related vulnerability to Aβ42 toxicity.

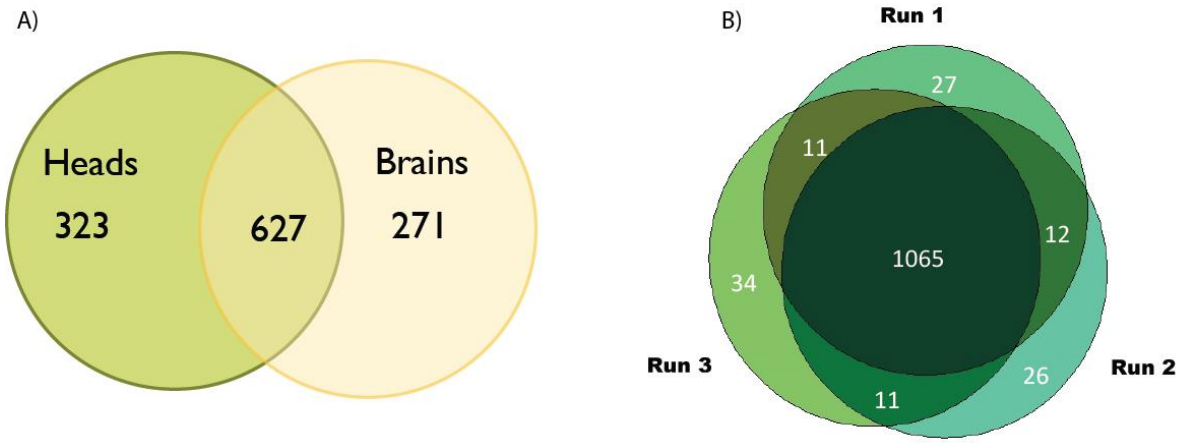


Figure 2.4 Results from the GeLC-MS^E analysis of *Drosophila* heads and brains. **A)** Overlap between proteins identified from technical replicate analysis of protein extracts from 10 heads vs protein extracts from 10 brains. **B)** Overlap between proteins identified from technical replicate analysis of brain samples

2.3.2 Experimental Design

We wished to identify proteins directly involved in age-related vulnerability to Aß42 toxicity. To achieve this we compared the brain proteomes of an adult-onset AD-*Drosophila* model, where Aß42 expression had been induced at two different ages, 5-12 days (young) and 20-27 days (old). Brains had been dissected 14 days post initial induction, a time point at which it would be expected for a climbing phenotype to be detectable. To allow us to probe Aß specific and age specific proteome responses in parallel, age matched controls were included where Aß42 expression had not been induced, **Figure 2.5A**. Prior to analysis by mass spectrometry levels of Aβ42 were quantified by enzyme-linked immunosorbent assay (ELISA). Data confirmed that both induced samples had similar levels of peptide, **Appendix B.** Biological repeat A was not included in the proteomic analysis due to higher levels of Aβ levels than the 4 other samples.

2.3.3 Proteomic Workflow

In order to obtain robust quantitative data, for each biological condition studied by our 1D-gel-LC-MS workflow, **Figure 2.5B**, four biological repeats were analysed, with each repeat being analysed by LC-MS in duplicate. Database searching was performed at a protein false discovery rate (FDR) of 5 %. Data were additionally filtered based on replication, with a protein being successfully identified if it was detected in a minimum of 2 biological replicates and three technical replicates. As instrumental noise is random in nature, true peptide signals should replicate, allowing the FDR to be reduced further.

2.3.4 The Mass Spectrometry-Based Proteome is Representative of the Genome

A common criticism of mass spectrometry-based proteomics is that methods used to prepare samples for analysis produce peptide solutions that are not fully representative of the original samples being interrogated. Due to the incompatibility of electrospray ionization (ESI) with high concentrations of detergent, protein extractions and digestions are commonly carried out at low detergent concentrations. Highly hydrophobic proteins such as membrane proteins are underrepresented as a direct result of not being soluble during proteolysis. Recently a method was introduced by JR Wisniewski et. al, that enables the use of high concentrations of detergent during sample lysis, prior to in-solution digestion enabling the solublisation and subsequent digestion of highly hydrophobic proteins prior to MS analysis [38]. This method however, requires larger amounts of starting material due to sample loss on molecular weight cut-off filters used for buffer exchange. If peptide level fractionation is performed-post digestion (commonly strong cation exchange) further sample loss would occur. Gel-based fractionation methods also facilitate the use of high concentrations of detergent, typically 4 % sodium dodecyl sulcate. Due to the low amount of starting material for the analysis, ~15 µg, it was hypothesized that the use of a gel-based method where fractionation is prior to digestion and extraction would limit possible sample loss by minimizing protocol steps, whilst solubilizing proteins in an unbiased manner.

In parallel to proteomic analysis, microarray mRNA transcriptional profiling was performed to examine gene expression profiles in relation to age and Aß expression. Microarray analysis was performed on

brain tissue and therefore provided the ideal benchmark for comparison with our proteomic dataset. All proteins identified from MS analysis, at a 5 % protein level FDR, and all genes detected in the

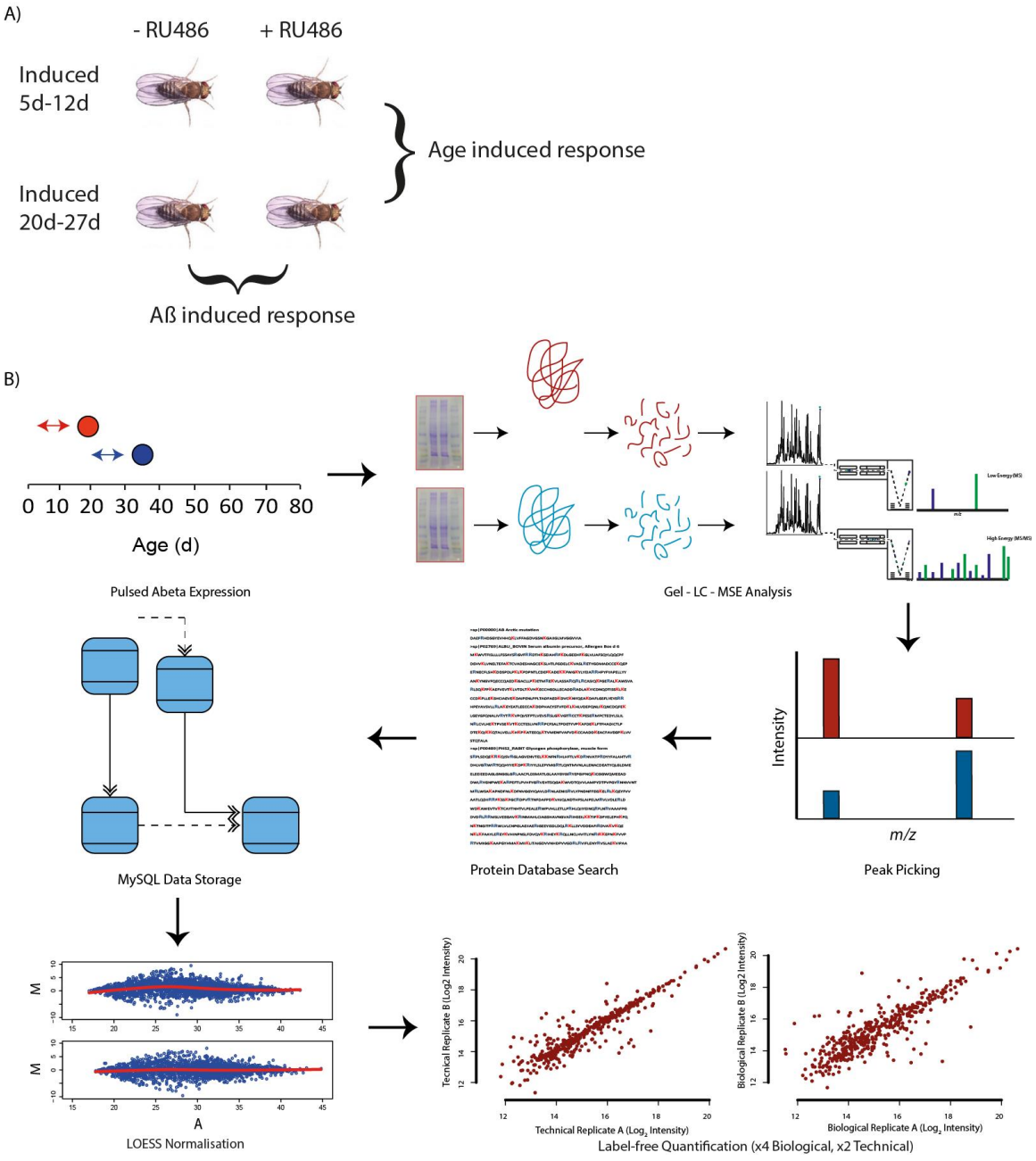


Figure 2.5 Experimental design and proteomic workflow for the unbiased screen of age-related vulnerability to Aβ toxicity. **A)** Experiment conditions examined. **B)** GeLC-MS^E workflow. Following arrows from top left, 15 Drosophila brains were homogenised in 30 ul SDS-PAGE loading buffer. Extracted proteins were separated by SDS-PAGE and each sample lane was separated into 10 bands. Bands were tryptically digested. Resultant peptide solutions were separated over an 80 minute gradient prior to analysis by MS^E on a Synapt HDMS. Database search results were stored in a custom built MySQL database

and peptide intensities were merged for gel bands common to a sample by summation based on protein accession. Data were normalized at the peptide level using an in-house written R script using locally weighted scatter point smoothing (LOESS). Relative quantification was achieved using the Hi3 technique.

microarray analysis were taken forward for gene ontology analysis. The two physiochemical properties, molecular weight and isoelectric point, were also calculated and compared between the two datasets. Gene ontology analysis and physiochemical property analysis was able to confirm that the two datasets compare well, **Figure 2.6**. The proportion of membrane proteins identified from mass spectrometric analysis agrees extremely well with that of the microarray results, confirming that the use of gel-based proteomic methods negates possible bias during the solubilisation process, **Figure 2.6A**. Our data suggests that the proteomic method preferentially identifies lower molecular weight proteins in comparison to the microarray results, **Figure 2.6D**. As proteins are proteolytically cleaved prior to analysis, protein size should only affect the number of possible peptide fragments that can be produced. This would suggest that MS analysis might preferentially identify larger proteins. Indeed it has been shown that quantification strategies, such as Hi3, can underrepresent the quantity of smaller proteins, or present a much larger measurement error [39]. It was hypothesized that the observed preferential identification of smaller proteins may be a result of the gel fractionation method, where proteins are separated based on their molecular weight. Better resolution is observed for proteins of a lower molecular weight, which would, therefore, aid their identification through decreased sample complexity. Testing our gel-based results against an in-solution method without any protein level fractionation, on the same samples, would be able to better explore this hypothesis.

2.3.5 Protein Quantification

Quantification was achieved using the previously described Hi3 technique. Hi3 is based on the observation that the summed intensity of the top 3 best ionizing peptides, as enzymatically derived from a protein, are directly proportional to the original protein concentration. A requirement for quantification is, therefore, the identification of a protein by three or more constituent peptides. A total of 1203 proteins were quantified, of all quantified proteins 478 proteins were common across all four conditions. Both technical and biological precision were found to be high, with a median Pearson's

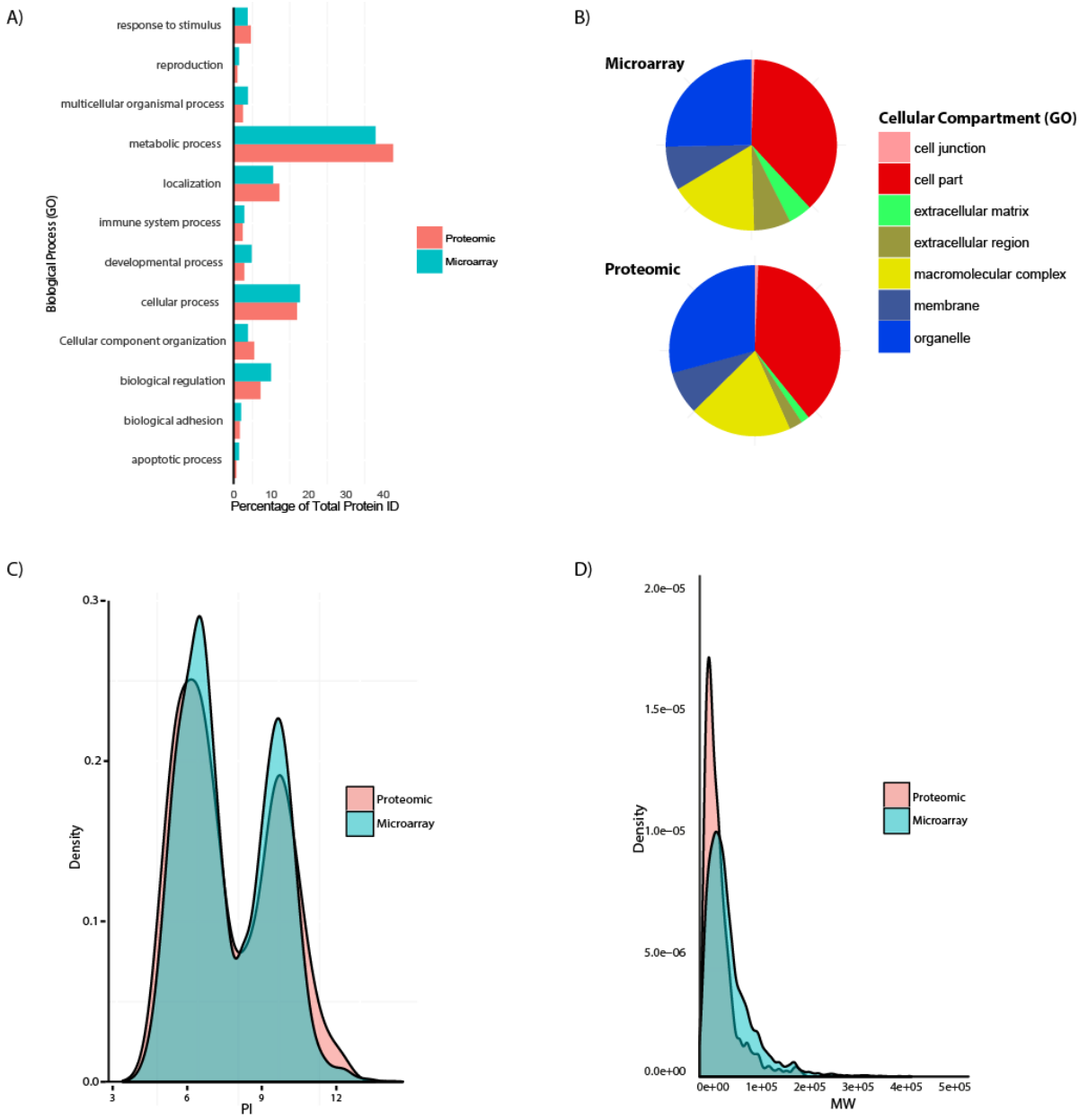


Figure 2.6 Comparison of proteomic and transcriptomic profiles by gene ontology and physiochemical properties. **A)** Comparison of proteomic and transciptomic datasets by gene ontology analysis; biological process terms. **B)** Comparison by cellular compartment terms. **C)** Kernel density plot of isoelectric point distributions. **D)** Kernel density plot of protein molecular weight distributions

correlation coefficient of 0.95 for technical repeats and 0.92 for biological repeats, representing a median coefficient of variation (CV) of 25.5% on repeat measurements. Based upon mean CV, power analysis indicated that four biological repeats would successfully identify fold changes of > 2 fold increase using a power of 0.8 as the threshold, **Figure 2.7**. Differential expression was based on calculated p values from each binary comparison and statistical power analysis that has been

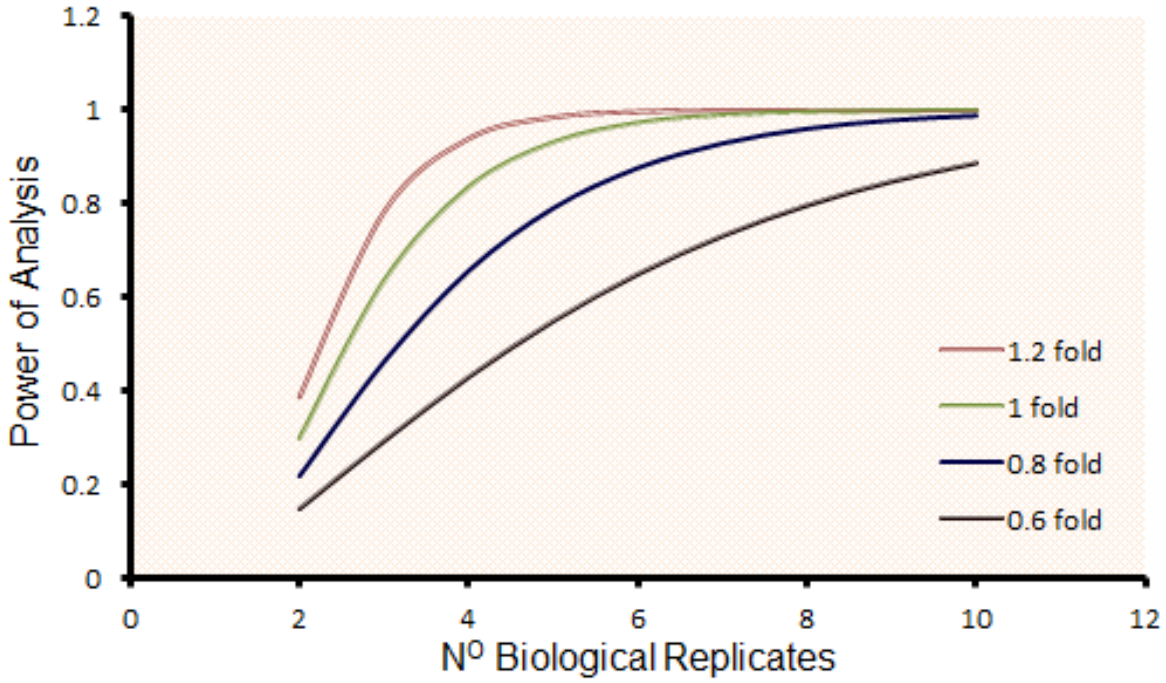


Figure 2.7 Power analysis. Power calculations were based on a mean coefficient of variance of 40 % and 2-10 biological repeats. Power of analysis has been calculated at four different predicted fold changes.

previously shown to provide an effective estimation of the appropriate effect-size cutoff (Cohen's *d*) as a function of protein intensity. Briefly, CV values of repeat quantitative measurements, of proteins identified within a condition, were plotted against their respective protein intensity, **Figure 2.8**. A rolling mean of CV values was plotted as a function of Hi3 protein intensity. This approach assessed global measurement error in reference to protein intensity. Interestingly our data did not display any observed reduction in measurement error with increasing protein intensity, a correlation which has been observed in previous label-free studies. This would suggest that the most significant source of error in our measurements was biological variability, which would not be affected by peptide intensity. As such we took forward the average of the rolling mean, 33 % CV, for effect size calculations. Proteins in each binary comparison that had an effect size larger than 1, indicating the mean of one group is found in the 85th percentile of that of the second group, were consider as displaying a large enough effect size abundance change to be taken forward for p value calculation. P values were calculated using a two tailed students t-test. Expression change was deemed significant at a p value <0.01, thereby limiting the number of false positives to less than one protein per comparison, based on the number of proteins tested. Results of the pairwise comparisons are supplied in **Appendix C.**

Differential protein expression was also defined based on a protein's absence from one condition, in a binary comparison, **Appendix D.** The rationale behind this approach is that if a protein is confidently identified in one condition, but is not detected in the other condition to which it is being compared, then an expression change must have resulted in a protein concentration below the limit of detection of our analysis. To reduce the number of false positive arising from these criteria, a protein was only differentially expressed if it met the criteria for protein identification in one condition and was completely undetected in all four biological repeats of the other condition.

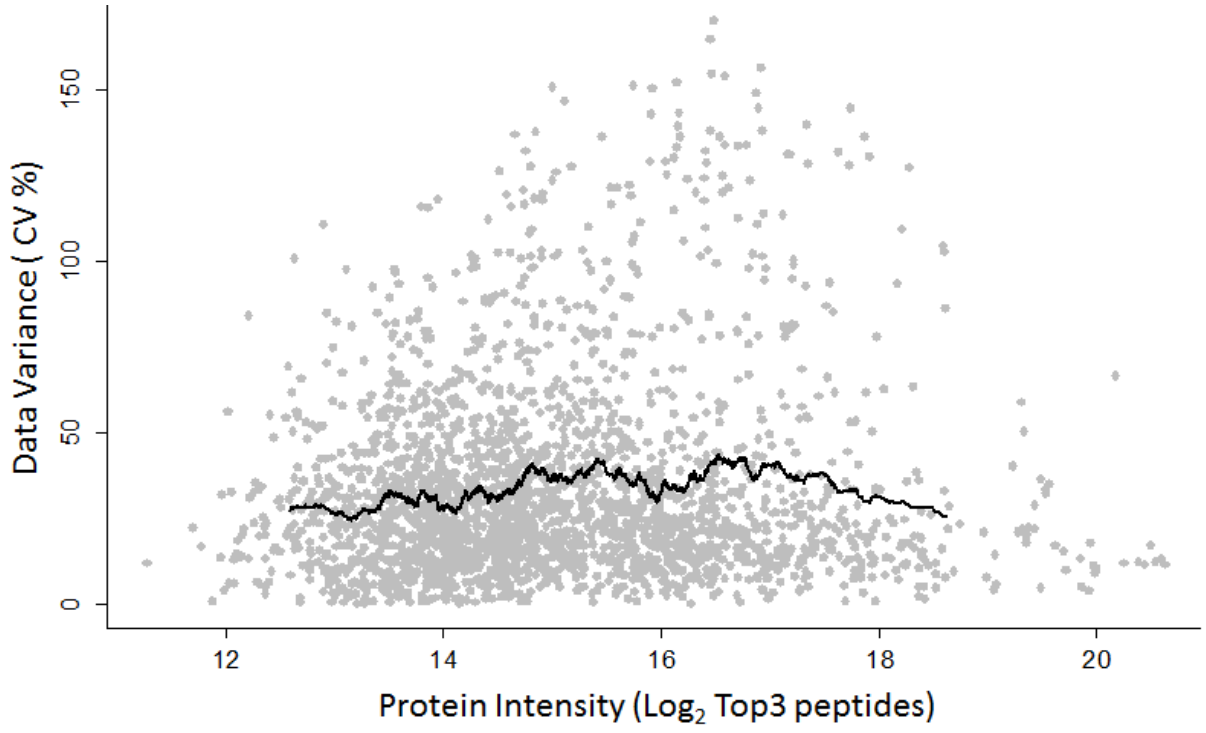


Figure 2.8 Estimation of global measurement error, as a function of Hi3 protein intensity, for effect size calculations. Data points represent measured data variance (% CV) of protein quantification (combined technical and biological replicates). The thick black line is a superimposed rolling mean

2.3.6 Age-Related Protein Expression Changes in the Presence of Aß42 Toxicity

We initially examined the relative protein expression levels in the brains of old Aß42 flies in comparison to young Aß42 flies. This pairwise comparison allowed the identification of 38 proteins that are a dysregulated as a result of age in the presence of Aß42 toxicity. These proteins represent potential candidates for contributing to age related vulnerability, **Table 2.1**. Gene ontology analysis of

these proteins did not identify any enriched GO terms for this subset of proteins, **Figure 2.9A**. Proteins involved in metabolic processes, a wide ranging term that encompasses both biosynthetic and catabolic processes were by far the most common. When compared to the background of all proteins identified in the screen, it is clear this does not represent true enrichment as a result of a cellular response but a tissue specific enrichment. Pathway analysis confirmed gene ontology results, showing a wide variety of functions for the 38 proteins but no particular enrichment of function, **Figure 2.9B**.

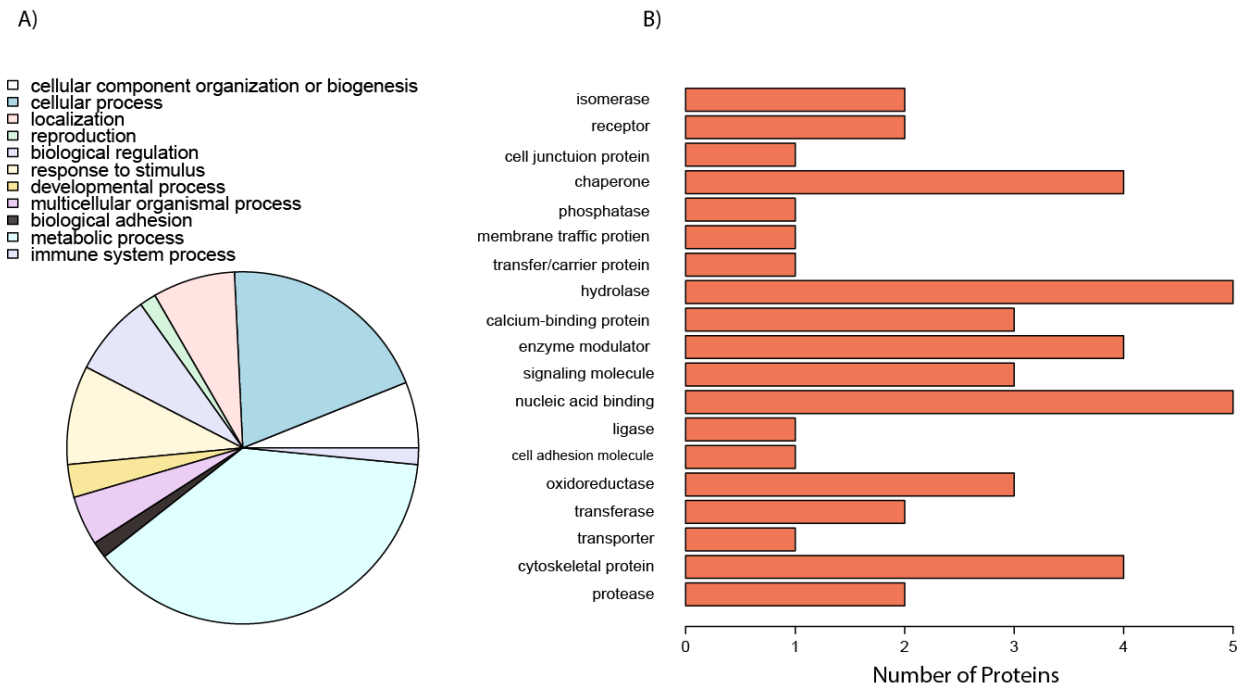


Figure 2.9 Gene ontology analysis results for differentially expressed proteins (Young Aβ42 vs Old Aβ42). **A)** Biological Process Terms. **B)** Panther Pathway Terms

Table 2.1 Differentially expressed proteins from the binary comparison, old + vs young + flies

Protein	Fold Change	Effect Size	P value
Glycyl-tRNA synthetase	1.531583695	1.24	0.0002925
L-lactate dehydrogenase	0.466199122	2.07	0.00747528
Clathrin-associated protein AP2 (AP50)	0.61110634	1.42	0.00929865
RH04549p	0.684319892	1.12	0.01020012
Peroxiredoxin 5	1.643725092	1.43	0.01078999
NAP1	Unique to Young Aβ		
14-3-3 protein epsilon	Unique to Young Aβ		
HSP27	Unique to Young Aβ		
Heterogeneous nuclear ribonucleoprotein A1	Unique to Young Aβ		
Microtubule-assocaited protein Jupiter	Unique to Young Aβ		
cAMP-dependent protein kinase type I	Unique to Young Aβ		
Eukaryotic translation initiation factor 3 subunit F-1	Unique to Young Aβ		
Heat shock protein 26	Unique to Young Aβ		
Heat shock 70 kDa protein cognate 2	Unique to Young Aβ		
Myosin regulatory light chain 2	Unique to Young Aβ		
CG9090	Unique to Young Aβ		
Proteasome subunit beta type	Unique to Young Aβ		
Eb1, isoform F	Unique to Young Aβ		
Sec22	Unique to Young Aβ		
Myofilin	Unique to Young Aβ		
Myosin heavy chain, muscle	Unique to Young Aβ		
Dihydroorotate dehydrogenase	Unique to Young Aβ		
Flightin	Unique to Young Aβ		
Troponin1	Unique to Young Aβ		
Tetraspanin 42Ef	Unique to Young Aβ		
CG11474	Unique to Young Aβ		
Sarcoplasmic calcium-binding protein 1	Unique to Young Aβ		
IA-2	Unique to Young Aβ		
cg991	Unique to Young Aβ		
40s ribosomal protein s5a	Unique to Young Aβ		
cg9512	Unique to Young Aβ		
Glycogen Synthase	Unique to Young Aβ		
cg12237	Unique to Old Aβ		
sd22712p	Unique to Old Aβ		
40s ribosomal protein s24	Unique to Old Aβ		
Putative peptidyl-prolyl cis-trans isomerase dodo	Unique to Old Aβ		
Acetylcholinesterase	Unique to Old Aβ		
Replication factor C subunit 3	Unique to Old Aβ		
Annexin B11	Unique to Old Aβ		

2.3.7 A Subset of Motor-Related Proteins are Upregulated in Young Aβ flies

Following gene ontology analysis results, differentially expressed proteins were search against known protein-protein interactions (PPIs) databases using the online tool, Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) [40]. Proteins were searched against a *Drosophila* specific database and only interactions with medium confidence were included in the output, **Figure 2.9**. Although the majority of proteins do not have any known interaction partners, two clusters of interest were identified. The main cluster contains six proteins that display large amounts of coordination with one another and a seventh protein, which interacts with a single protein within the main cluster. To test for the functional significance of these clusters enrichment analysis was conducted within STRING, to test for PPIs that have common biological functions, molecular functions and/or cellular components. After correction for multiple hypothesis testing no molecular or biological functions were enriched, however a number of cellular components categories reached significance ($p < 0.05$), **Table 2.2**. A common theme among the enriched categories is that they are integral to cytoskeleton structure and function, indeed cytoskeleton part is found to be significantly enriched ($p = 0.003$). Linking this data with the interaction network, **Figure 2.10,** it is found that the main protein cluster are all classified as a cytoskeleton component. Interestingly analysis of expression profiles confirms that proteins within this cluster display similar expression levels. All proteins were highlighted from their confident detection in samples from young +RU flies but their absence in samples derived from old –RU flies. Further examination of the data from the screen determined that they are in fact almost exclusively identified in young +RU flies. These results suggest a unique and coordinated response to amyloid expression that changes with age.

Examination of the 7 protein cluster identified Wings up A (WupA) as the protein to coordinate with the highest number of partners. It was found to directly interact with five other proteins within the interaction cluster and therefore makes an interesting target to explore. WupA encodes Troponin I, a key protein in muscle cells that regulates the sliding of thin over thick filaments, but more interestingly is thought to be involved in the development and maintenance of the nervous system [41], [42]. Troponin I, is well characterized in patients with heart disease, where it is a well-known stress marker but there is little information on its function within the brain [43]. Second to WupA, in terms of number of interacting partners is myosin regulatory light chain (RLC), for which there is more known about its

function within the central nervous system. Myosin RLC is normally associated with the neck region of the myosin II protein complex and is best characterized in its role in the regulation of myosin-based contraction in muscle cells. Myosin II is also found post-synaptically and has been demonstrated to directly interact with NMDA receptors [44]. Importantly the myosin II complex has been found to be directly involved in mediating the tau toxicity mechanism, through this interaction [45].

Table 2.2 Enrichment analysis for cellular components from STRING analysis. Proteins included were identified as differentially expressed (old + vs young +)

Cellular Compartment	N° Genes	P value	P value_fdr
muscle myosin complex	3	4.90E-07	4.80E-04
myosin II complex	3	4.10E-06	2.00E-03
myofibril	4	1.10E-05	2.70E-03
contractile fiber part	4	1.30E-05	2.70E-03
contractile fiber	4	1.40E-05	2.70E-03
intracellular part	28	2.20E-05	3.50E-03
cytoskeletal part	10	2.50E-05	3.50E-03
intracellular	28	3.40E-05	4.20E-03
cytoskeleton	10	5.10E-05	5.60E-03
cytoplasm	20	8.50E-05	8.40E-03
cytoplasmic part	17	1.00E-04	9.00E-03
cell	29	1.30E-04	9.90E-03
cell part	29	1.30E-04	9.90E-03
myosin complex	3	1.50E-04	1.10E-02
myofilament	2	2.90E-04	1.90E-02
sarcomere	3	3.20E-04	2.00E-02
intracellular organelle	24	3.90E-04	2.20E-02
organelle	24	5.00E-04	2.70E-02
actin cytoskeleton	4	6.80E-04	3.50E-02
macromolecular complex	18	7.00E-04	3.50E-02
A band	2	7.40E-04	3.50E-02
non-membrane-bounded organelle	13	8.10E-04	3.50E-02
intracellular non-membrane-bounded organelle	13	8.10E-04	3.50E-02
intracellular organelle part	17	9.00E-04	3.70E-02
organelle part	17	1.10E-03	4.20E-02

Models have shown tau to be essential for Aβ-induced NMDA dependent postsynaptic dysfunction [46]. In *Drosophila* loss of function mutants, for Zipper and spaghetti squash, the genes encoding heavy chain (HC) and RLC of myosin, display increased tau toxicity. These results suggest an age-

related loss of myosin RLC in the presence of Aβ that could be directly involved in the increased toxicity observed in older flies.

Little is known about the function of many of the seven proteins' functions within the brain and the results from the screen could represent an interesting avenue of research that has little to no focus currently in the AD research field. It is possible that the identified protein network could be closely linked to myosin RLC toxicity mediation; however this is based purely upon PPI data and requires significant validation. Prior to this, validation of the results from the screen for this subset of data by an orthogonal quantative/semi-quantitative method should be carried out.

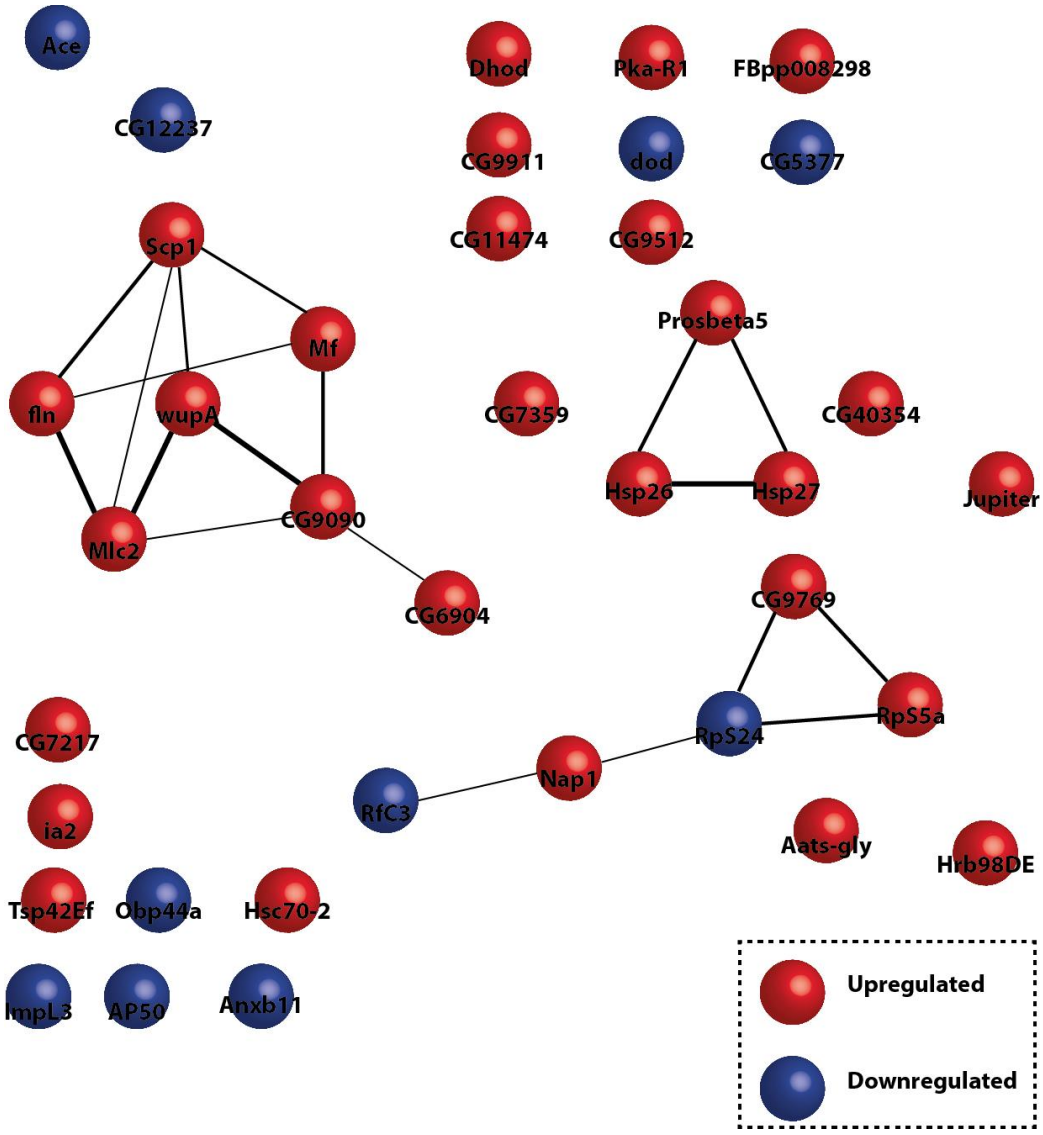


Figure 2.10 Annotated output from the protein-protein interaction analysis of 38 proteins identified as differentially expressed in the binary comparison of old + flies vs young + flies. Analysis carried out using STRING. Lines indicate a

protein-protein interaction and thickness of lines is representative of the confidence. Only PPIs with medium or higher confidence are included in the output

2.2.8 Heat Shock Proteins Show a Mixed Response to Aβ Expression and Age

Pathway analysis of the results from the old Aβ42 flies vs young Aβ42 flies binary comparison identified expression changes for 4 chaperone proteins that are known to directly interact. Dysregulated chaperone expression is an expected cellular response to the increase in misfolded protein load caused by Aβ42 aggregation. Interestingly these proteins show an age-related response in addition to an Aβ42 response. This led us to further investigate the expression profiles of the 4 proteins across all tested conditions. Three of the chaperone proteins were found to be heat shock proteins (Hsp), namely Hsc 70, Hsp 27 and Hsp 26, with the fourth protein being the third isoform of 14-3-3 protein epsilon. Heat shock proteins are a set of highly conserved molecular chaperones that are induced in response to a variety of physiological and environmental stresses, allowing cells to survive potentially lethal conditions. When acting as molecular chaperones, Hsps bind to misfolded proteins and catalyze, often in an ATP-dependent manner, proper re-folding of proteins to prevent potential aggregation. The family of Hsp 70 proteins' main function is to assist in the folding of nascent polypeptides under normal conditions. The chaperone family additionally aids the assembly of protein complexes and the transport of proteins across membranes. They act by holding translated or nascent peptide chains in a state suitable for folding when released into the medium. Hsp 70 is commonly observed to be upregulated in disease states as studied by proteomics. It can be seen that as predicted, Hsp70 is upregulated in response to Aß expression, **Figure 2.11A**. Interestingly the data shows a small reduction in expression in response to Aß, with age, however for the conditions tested this did not reach significance ($p = 0.11$). This is in contrast to expression levels quantified in the two non-induced controls, with Hsp70 expression elevated ($p= 0.038$) in the older flies. This is in contrast to previous studies where global levels of Hsp70 during *Drosophila* ageing have been observed to reduce with increasing age, an observation that has been hypothesised to be an explanation for reduced stress resistance in aged *Drosophila* [47]. Our data, however, detects significant alterations in heat shock 70 kDa protein cognate 3 (Hsc70-3) levels but no other members of the Hsp70 family. The data therefore indicates a more specific cellular response than previous publications have monitored.

Hsc70-3 is the functional counterpart of binding immunoglobulin protein (BiP) the endoplasmic reticulum (ER) resident chaperone that binds newly synthesised proteins as they are translocated into the ER. If misfolded protein load increases BiP is a key component of the ER associated degradation response (ERAD). Increased levels of BiP following the induction of a variant of Aβ42 that is known to rapidly aggregate could, therefore, be indicative of the activation of ER stress pathways in response to increased levels of misfolded protein.

The Hsp70 chaperone family's association with neurodegenerative diseases has been previous examined and the overexpression in a polyglutamine disease and an AD model has been demonstrated to successfully suppress protein aggregation and ameliorate the corresponding disease phenotypes [48]. Our results potentially implicate Hsc70-3 levels upon ageing with increased vulnerability to Aβ peptide.

Hsp26 and Hsp27 are small heat shock proteins (sHsp), have distinct roles in response to stress but share 48% sequence identity. sHsps are characterised by the presence of a highly conserved "α crystallin" domain. Functionally sHsps bind to partially unfolded proteins, thereby stabilising the structure rather than actively refolding proteins themselves. Hsp26 is known in yeast, upon temperature change, to switch from an inactive state to an active chaperoning state. In addition to chaperoning activity Hsp27 protects against cell death, which can be induced by various stimuli including oxidative stress, a well-documented stage of the amyloid cascade hypothesis. Interestingly Hsp27 has been shown to reduce the production of reactive oxygen species (ROS) caused by tumour necrosis factor- α (TNF- α), a protein that has long been linked with the inflammatory response in AD and is a proposed therapeutic target [49], [50]. The potential of Hsp27 as a potential therapeutic target in AD is enhanced by its known ability to prevent apoptosis through its protective function against cellular stressors including heat and reactive oxygen species [51]. Hsp27 has been shown to inactivate Bax, a pro-apoptotic molecule, in addition to its well characterised interaction with cytochrome c, through which Hsp27 can negatively regulate cell death [52].

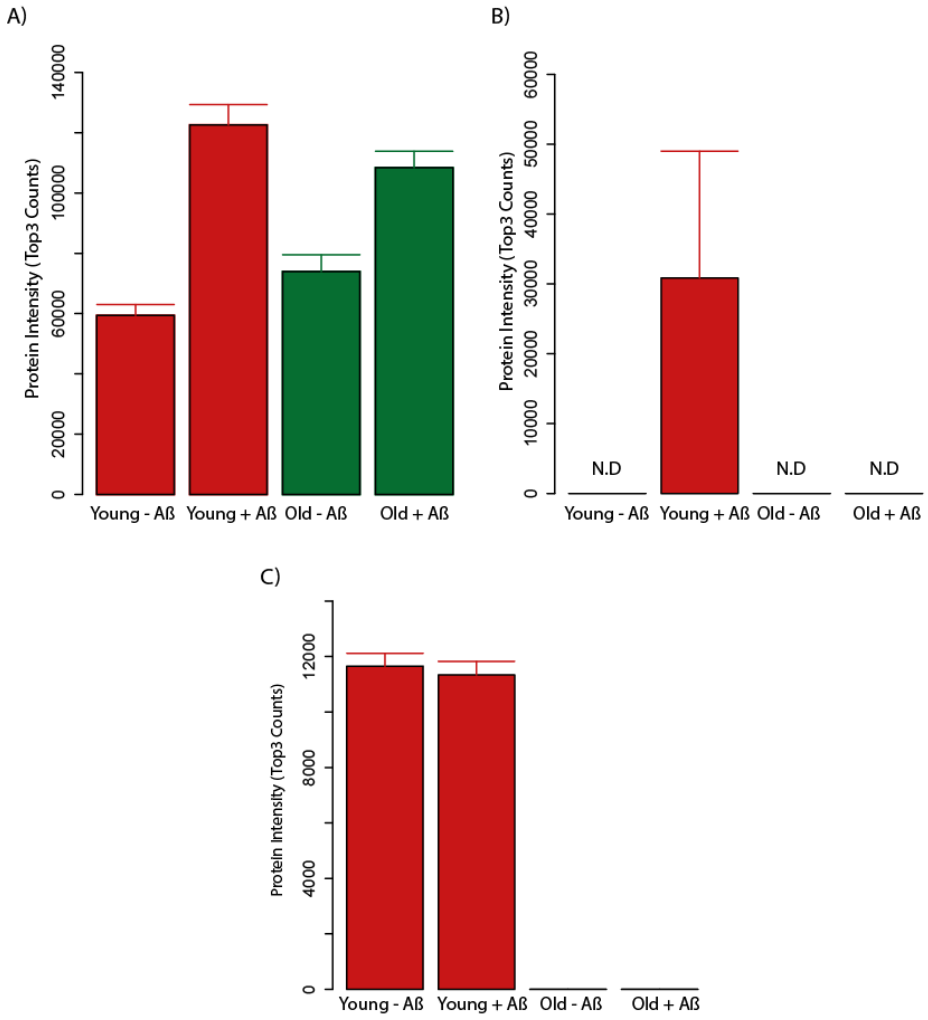


Figure 2.11 Relative protein expression levels for three differentially expressed heat shock proteins **A)** Expression levels for Hsc70-3. B) Hsp 26. C) Hsp 27. Error bars represent standard error of mean (SEM)

Our proteomic results quantify both an age and Aβ expression response from Hsp26 and 27, however their relative expression levels are distinct from one another, **Figure 2.11B&C**. Hsp26 is identified in only one of the four conditions, young Aβ flies. These data would suggest ageing reduces the level of Hsp26 in response to aggregation of Aβ42. Based on the known function of Hsp26 it is likely this would have a deleterious effect on cellular function. Unfortunately due to the limits of detection of our method, we were unable to ascertain the effects of ageing on the expression of the protein under normal physiological conditions. In contrast Hsp27 does not show any detectable response to Aβ but expression levels are dramatically reduced with age. In the samples obtained from old *Drosophila,* due to both conditions being below the limit of detection, nothing is known about expression in response to Aβ. Apoptosis is thought to be a major pathway to neuronal death in AD and a reduction

in basal expression levels of Hsp27, could cause an increase in apoptotic cell death in old Aβ flies. Further to this, it has been previous demonstrated that overexpression of Hsp27 successfully ameliorates AD symptoms in an APP/PSI mouse model.[53] Behavioral assays found learning and memory to be significantly improved upon Hsp27 overexpression, which was attributed to a reduction in amyloid deposits in both the hippocampus and cortex regions of the mice.

Small heat shock proteins have also been shown to have beneficial effects in wild-type organisms. Overexpression of either Hsp26 or Hsp27 has been shown to extend *Drosophila* lifespan by up to 30 %.[54] As expected there was an increased resistance to paraquat, along with an observed reduction in fecundity. Importantly lifespan extending effects are still present when neuronal specific expression is undertaken [55]. Interestingly the extent of extension was, as in the previous study, found to be around 30% suggesting that all or the majority of the observed lifespan extension can be attributed to better control of neuronal homeostasis. The same study also found Hsp27, but not Hsp26, to attenuate mild polyglutamine-induced toxicity and reduce parkinsonism movement disorder as studied in a *Drosophila* model.

2.3.9 GTPase-Related Protein Expression Changes are Enriched in Old Aβ *Drosophila*

Results from gene ontology analysis and PPI analysis of the pairwise comparison of old control flies vs old Aβ flies revealed a number of GTPase-related proteins were quantified with statistically significant differences. Enrichment analysis for biological processes based on identified PPIs found small GTPase mediated signal transduction to be most enriched, **Table 2.3**. Functional analysis of these proteins found a number to belong to the Rab family. Rab proteins are the largest family of small Ras-like GTPases with 29 predicted Rabs in *Drosophila* and over 70 Rab or Rab-like proteins thought to be encoded in humans. Highly conserved across species, the Rab family serves a number of biological functions, but are best known for their involvement in membrane trafficking circuits [56]. Such proteins are present in all compartments of cell, including nucleus, mitochondria, plasma membrane, endoplasmic reticulum and Golgi. Many Rab proteins have been well characterised for their role in endocytic and exocytic membrane trafficking which is integral to proper synaptic function [57]–[59]. Rab proteins are able to cycle between membrane bound and cytosolic depending on their

nucleotide bound state. When GDP-bound, also referred to as “inactive”, rab proteins insert to the respective compartment membrane. Upon conversion into the “active”, GTP-bound state interactions with effector proteins involved in trafficking pathways are possible.

It is well known that for proper neuronal function, specialized intracellular membrane trafficking must be maintained. This is especially true at synaptic junctions. It was recently reported that half of all *Drosophila* Rabs function within neurons and that these neuronal Rabs are predominately active at synapses where they mark synaptic recycling of endosomal compartments.[60] With Rab proteins so closely linked to the tight control of synaptic membrane trafficking it could be hypothesized that the observed abundance alterations could potentially have direct effects on neuronal function. Analysis of our proteomic results has revealed seven of these GTPase related proteins display significant protein expression changes both in response to Aβ and age, **Figure 2.12**. Interestingly, although some common expression patterns are present, particularly between Rab 10, Rab 30 and Ran, expression responses have varied profiles.

One notable protein, Rab3, is highly conserved across species and in humans, consists of four isoforms (A-D), all of which are enriched within the brain and in particular synaptic vesicles. The subfamily have a well characterized functional role in the endocytic pathway and account for >25% of total rab GTP-binding activity. *Drosophila* only express one form of Rab3; Rab3A and is of particular interest as our proteomic results reveal expression changes linked to both Aβ and ageing phenotypes. It can be seen from **Figure 2.12** that no expression changes are observed across the two young fly conditions (+/- RU), however older control flies (-RU) show a significant upregulation ($p < 0.05$), which is reduced upon Aβ induction. Although Rab3A has already been implicated in synaptic dysfunction in late stage AD patients, through the reduction in expression levels compared to healthy controls, the functional consequence of this loss in AD has not yet been studied [61]. Rab3A’s link with age-related vulnerability to AD is currently unknown. Interestingly it has been proven in *Drosophila* that Rab3A plays a key role in regulating the distribution of presynaptic components to active zones [62]. In Rab3A mutants the resultant phenotype was more localized active zones, with higher release probabilities. Knockouts of the Rab3 subfamily in different model organisms, however have been found to have

Table 2.3 Enrichment analysis for biological process terms from STRING analysis. Proteins included were identified as differentially expressed (old + vs old-)

Biological Process Term	No Genes	P value	fdr
small GTPase mediated signal transduction	6	1.80E-06	8.50E-03
intracellular signal transduction	8	3.20E-06	8.50E-03
Ras protein signal transduction	5	4.40E-06	8.50E-03
cytoskeleton organization	10	1.10E-05	1.60E-02
single-organism cellular process	27	1.70E-05	2.00E-02
microtubule cytoskeleton organization	8	2.30E-05	2.20E-02
cellular process	29	2.80E-05	2.40E-02
establishment of protein localization	7	3.60E-05	2.70E-02
microtubule-based process	8	1.30E-04	8.10E-02
organelle organization	13	1.50E-04	8.10E-02
Rab protein signal transduction	3	1.50E-04	8.10E-02
protein transport	6	2.70E-04	1.30E-01
response to stimulus	17	3.00E-04	1.30E-01
macromolecule localization	8	3.60E-04	1.40E-01
cellular component organization	16	3.60E-04	1.40E-01
protein localization	7	4.30E-04	1.40E-01
organic substance transport	7	4.60E-04	1.40E-01
single organism signaling	12	4.90E-04	1.40E-01
axis specification	5	5.40E-04	1.50E-01
cell communication	12	6.70E-04	1.80E-01
spindle organization	5	1.00E-03	2.50E-01
single-organism process	28	1.10E-03	2.60E-01
cellular calcium ion homeostasis	2	1.50E-03	3.40E-01
regulation of biological quality	9	1.60E-03	3.40E-01
calcium ion homeostasis	2	1.60E-03	3.40E-01
cellular divalent inorganic cation homeostasis	2	1.60E-03	3.40E-01
signal transduction	9	1.80E-03	3.50E-01
single-organism organelle organization	9	1.90E-03	3.70E-01
oocyte microtubule cytoskeleton polarization	2	2.00E-03	3.80E-01
divalent inorganic cation homeostasis	2	2.40E-03	4.40E-01
cell cycle process	7	2.80E-03	4.90E-01
biological regulation	18	3.30E-03	5.70E-01
RNA interference	2	3.60E-03	6.00E-01
mitotic spindle organization	4	4.20E-03	6.50E-01
protein targeting to mitochondrion	2	4.40E-03	6.50E-01
protein localization to mitochondrion	2	4.40E-03	6.50E-01
cellular localization	7	4.70E-03	6.80E-01
endomembrane system organization	3	4.90E-03	6.90E-01
cellular response to stimulus	10	4.90E-03	6.90E-01

more serious effects. Mice knockouts for all four Rab3 isoforms have been shown to be unviable as they die during development, demonstrating a key role for the Rab3 subfamily in neuronal function [63].

The expression profile of the GTPase related proteins, **Figure 2.12**, also identifies three proteins that have similar quantification profiles both in terms of the measured response across the four conditions and absolute expression levels. The proteins in question, Rab 10, Rab 30 and Ras-related nuclear protein (Ran) have different known membrane trafficking roles, however, our data indicates their response to be coordinated. Rab 10 is localized in the Golgi and ER and is known to function in Golgi to basolateral membrane transport [64]. Rab 30 is also Golgi localized and is required for the integrity of the apparatus morphology itself [65]. Rab10 has been identified by a previous proteomic study of AD patients to be down regulated in post synaptic density (PSD), however its role in the disease has not been studied further [66]. Rab30 has not been previously reported in AD studies and represents a novel finding. Ran again functions as a molecular switch to control cellular transport and mitotic spindle assembly amongst others. High levels of Ran have been directly correlated with human tumourigenesis, however the reverse is seen in cases of AD [67]. Similar to Rab10, although downregulation has been seen in AD patients, the cause and subsequent consequence of this change is unknown. All three of the proteins display downregulation in the presence of Aβ, which agrees well with current literature. The quantification of EndoA, shows an expression profile that is different to the other 6 GTPase related proteins. The protein is only identified in the old Aβ42 flies and with a high total intensity. These data indicate a large upregulation upon Aβ induction but only for aged flies. EndoA, therefore, represent an interesting target for affecting age-related vulnerability to Aβ42. Interestingly Endophilin I (a homologue of EndoA) has been previously linked to AD, in both mice and humans. A study by Y. Ren and co-workers initially used quantitative proteomics to identify increased levels of Endophilin I in transgenic mice that overexpress Aβ [68]. These findings were the validated in humans with increased levels in AD patients confirmed by western blot quantification. The functional significance of this increase was subsequently investigated and was found to induce neuronal death through the increase of the stress kinase c-Jun N-terminal kinase. Our data expands upon these previous findings by linking these cellular alterations to ageing in addition to Aβ42 levels.

2.2.10 The Proteome and Transcriptome Show Low Correlation

Numerous studies have investigated the correlation between mRNA transcription and protein abundance. Early studies, often looking at simple organisms such as yeast found low levels of correlation between protein abundance and mRNA levels (R^2 = 0.3-0.6) [68]. These studies were limited by the available proteomic technologies that could only quantify hundreds of proteins unless extensive fractionation was employed. It was hard to discern from such studies whether the low correlations reported were indeed providing evidence for significant amounts of pre and post translational controls or were the result of instrumental bias/limitations. Indeed it is known that a large number of post-translation mechanisms exist that influence protein turnover rates and ultimately abundance. These mechanisms include proteasomal degradation and endocytic lysosomal systems. Phosphorylation, a highly dynamic post translational modification, is known to modulate cell division and differentiation. MicroRNA's (miRNA) on the other hand, represent pre translational

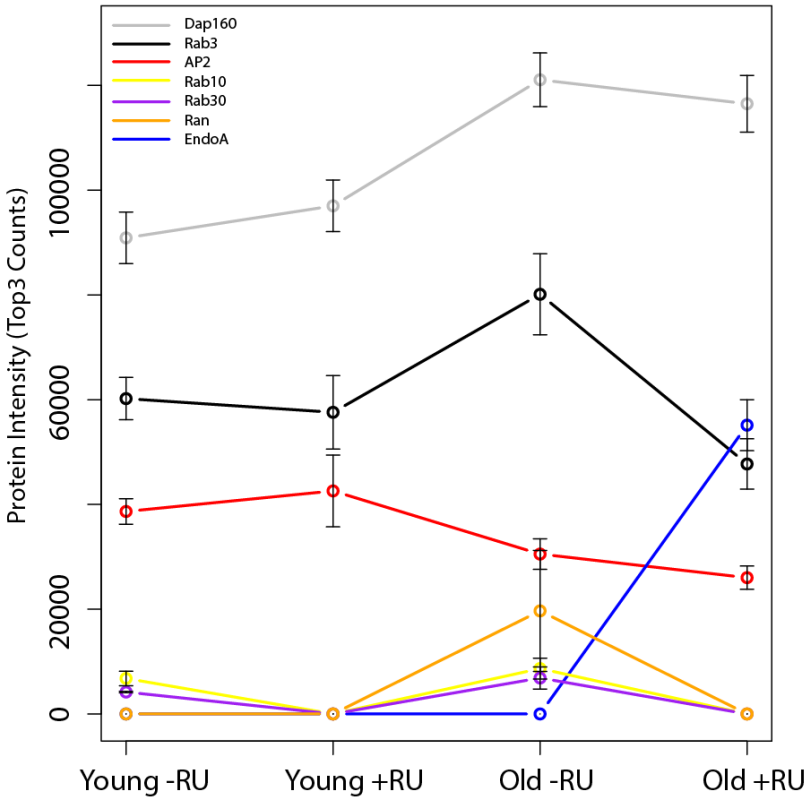


Figure 2.12 Expression levels for 7 small GTPase related proteins that display differential expression with respect to age. Expression levels with 0 counts is indicative of the protein not being successfully quantified. Error bars represent SEM.

control mechanisms. These endogenous short RNAs (~22 nucleotides) act as a class of regulatory elements to dampen the expression of protein coding genes [70]. It has also been demonstrated that miRNAs also reduce the levels of many targeted mRNAs [71]. What is not known is how much each of these elements contributes to final protein abundances.

Improvements in proteomic technologies have dramatically increased proteome coverage, however the majority of studies still report a correlation of protein and mRNA levels in the region of $R^2 = 0.6$, suggesting that just over half of protein level variability can be explained by the regulation of mRNA abundance.

Alongside proteomic analysis microarray analysis was conducted on all four conditions to identify differentially regulated mRNA levels. Samples were dissected at exactly the same time as those taken forward for proteomic analysis. The data therefore represents the ideal experimental design for determining the correlation between mRNA and protein levels in *Drosophila* brain tissue. For calculation of protein-mRNA correlation, the average protein Hi3 intensity across replicates was correlated with average microarray fluorescence, both plotted on $\log_2$ scale. Unfortunately proteome coverage achieved in this study was significantly lower than the number of mRNAs quantified by microarray analysis. To improve protein coverage and with the aim of reducing analysis bias, protein-mRNA levels were calculated at a 5 % protein FDR level (no filtering for replication). Spearman's rank correlations were calculated for each condition individually and when combined, **Figure 2.13.**

Protein to mRNA correlation was found to be low for each condition, **Figure 2.13A** and for the combined dataset, **Figure 2.13B**. Average Spearman's rank correlation across conditions was low at 0.28. Calculated correlations are indeed lower than those presented in many other recent proteomic studies. A major difference, however, between the majority of publications presenting mRNA to protein abundance correlations and our current study is that quantitative data is derived from a complex tissue rather a homogenous cell line. Tissue is a heterogeneous population of different cells types which for *Drosophila* brain tissue includes neurons, cortex glia and surface glia. It is known that different cell types will have inherently different mRNA to protein translation rates. mRNA to protein ratio calculations for tissue samples will therefore be an average of all these rates, leading to reduced overall correlation. Adding weight to this hypothesis are data from two research labs who have

recently presented the first drafts of the entire human proteome as defined by mass spectrometric analysis [72], [73]. The data from M. Wilhelm and co-workers has been compiled from a number of different laboratories and consist of a total of 16,857 LC-MS/MS experiments yielding 18,097 protein identifications. In the process proteomic results from a number of different tissues were compiled. This allowed the authors to determine mRNA-protein correlations for 12 different human tissues. Spearman's rank correlations ranged from 0.31 up to 0.56 for the different tissues which the authors acknowledged was lower than expected and agrees with the data presented here. Interestingly, despite the low correlation and the high variation of protein to mRNA levels across different tissues, mRNA-protein ratio was found to be consistent across tissue types. It would be interesting to determine if this held true in model organisms such as *Drosophila,* this, however, is beyond the scope of our experimental data.

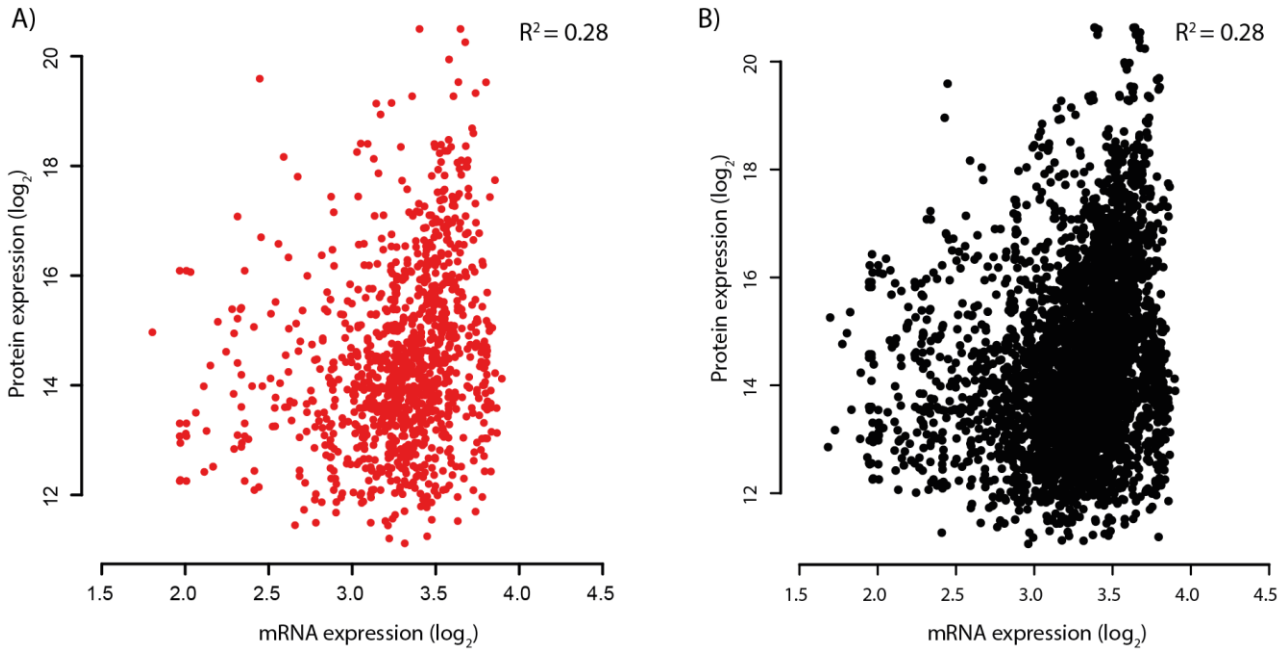


Figure 2.13 mRNA and protein levels show low correlation in *Drosophila* brain tissue. mRNA levels based on microarray florescent level and protein levels based on the Hi3 method. **A)** mRNA- protein level correlation plot for a single condition, Young +Aβ. **B)** mRNA- protein level correlation plot using results combined from the four different conditions.

Validation of Quantitative Results

Despite vast improvements in mass spectrometry-based proteomics quantitative accuracy, the most common approach to validate discovery derived results is still the semi-quantitative immunoassay technique, Western blotting. There is now a move towards using targeted mass spectrometry analysis workflows, such as SRM, to validate previous discovery results. Unfortunately the lab was not equipped with instruments that were able to perform targeted analysis and we therefore attempted to validate a small number of our statistically significant results with immunoassay quantitation. Antibodies for Rab-3, Ran, EndoA, AP50 and Hsc-70 were either obtained from other labs or where available purchased.

Western blot quantification of protein levels in *Drosophila* brain is almost exclusively carried out on head homogenate due to the number of brains required for detection by Western blot. We therefore tried to validate a set of our quantitative results by repeating the same experimental design as used previously, **Figure 2.5A** and subsequently quantifying protein levels from head homogenate extracts, **Figure 2.14**. Pleasingly similar expression profiles were observed for number of the proteins tested including Ran, AP50 and Rab3. Unfortunately only one statistical comparison reached significance, mostly likely due to a combination of the lower levels of precision offered by Western blotting, in comparison to quantitative mass spectrometry, and the increased sample complexity when analyzing head homogenate instead of brain.

Data for Hsc70-3, **Figure 2.14E**, were only collected for the young + & - as these two conditions were found to have the largest difference in expression levels, based on the proteomics data. Again similar expression alterations were observed for the blotting results when compared to those from the proteomics, with a slight increase as a result of Aβ42 expression, however the difference is not statistically significant. For all western blots a total of four biological repeats were analysed, and, based on the resulting data, this does not appear to be sufficient to provide the power to detect protein abundance alterations as confidently quantified by mass spectrometry. It is likely that if further biological repeats were carried out results would reach statistical significance. This is especially true for Ran, AP50 and Hsc70-3 were measured protein levels were found to be fairly consistent.

Results from EndoA blotting, **Figure 2.14C** found very little change in protein abundance across the four conditions in comparison to the mass spectrometry derived results. EndoA, in addition to playing a key role in synaptic endocytosis has also been shown to localise the photoreceptors of eye imaginal disc [74] Based on this information, the results, which do not demonstrate what is recorded by mass

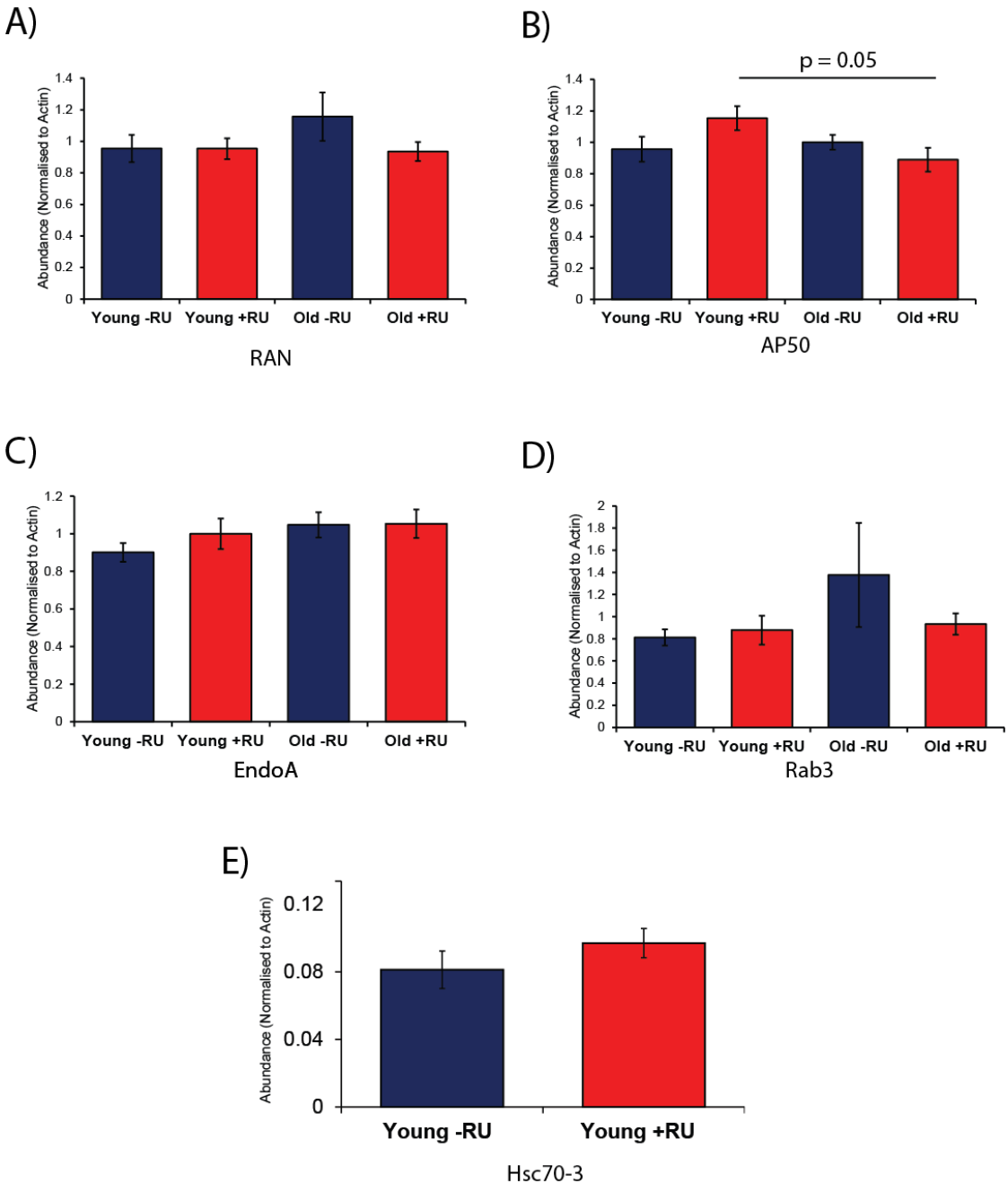


Figure 2.14 Validation of proteomic results by Western blotting analysis. Results shown are levels of A) Ran B) Clathrin associated protein AP2 (AP50) C) Endophillin A D) Rab3 and E) Heat shock cognate 70 isoform 3 normalised to levels of actin. For all analysis n=4. Statistical testing was by student t-test with unequal variance.

spectrometry, might be a result of expression levels in the eye effecting the quantitation. Overall the western blotting results, although promising, demonstrate the difficulty in validating *in vivo* results by immunoassay blotting. Ideally results could be validated by SRM or PRM analysis however this is currently beyond the abilities of the lab due to available instrumentation.

2.4 Conclusion

Ageing is the major risk factor for Alzheimer's disease, however the correlation between age and AD prevalence is not well understood. With an ageing population it is important that the association between age and neurodegeneration is studied in more detail. Here a label-free mass spectrometry-based quantitative proteomic approach was used to successfully reveal a set of proteins that could potentially be implicated in the age-related vulnerability to Aβ42 expression, as has been previously published, in an adult onset *Drosophila* AD model. Three heat shock proteins, Hsp70, Hsp26 and Hsp27 were identified as differentially expressed from the screen. Hsp26 and 27 displayed robust expression changes as a result of ageing whilst Hsp70 showed differential regulation in response to Aβ expression and slight but not statistically significant downregulation with age in the presence of Aβ. Interestingly HSP70 over-expression has been previously shown to ameliorate symptoms in AD models but has not been linked to ageing in AD patients. These findings suggest that age-related changes in proteostasis, in particular chaperone activity, may well contribute to the close link between ageing and late-onset Alzheimer's disease. Enrichment analysis of differential expressed proteins from the binary comparison of old - RU flies vs old Aβ flies established a set of seven GTPase-related proteins for follow up validation. Closely linked to endo/exocytosis, these proteins, though their diverse roles in membrane trafficking, help to maintain proper synaptic function. Their potential dysregulation upon ageing and Aβ expression could, therefore, have significant implications on brain function. Although western blotting was not able to confidently validate the mass spectrometry data, most likely due to the lower levels of precision provided by the immunoassay technique and an inability to analyse brain tissue, a number of the proteins tested had expression profiles that correlated well. This, taken together with the consensus of our data and previous literature, provides confidence that our results are truly representative of age-related vulnerability to Aβ42 aggregation.

Transgenic lines for all of the seven GTPase-related proteins have been obtained to conduct follow up research. Flies have been backcrossed into the w118 background to ensure a homogenesis stock genetic background. A lifespan screen is intended to be carried out to allow the identification of genetic alterations that extend lifespan in the presence of Aβ42 expression.

2.5 References

[1] R. Brookmeyer, E. Johnson, K. Ziegler-Graham, and H. M. Arrighi, "Forecasting the global burden of Alzheimer's disease.," *Alzheimers. Dement.*, vol. 3, no. 3, pp. 186–91, 2007.

[2] J. Hardy and D. J. Selkoe, "The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics.," *Science*, vol. 297, no. 5580, pp. 353–6, 2002.

[3] C. L. Masters, G. Simms, N. A. Weinman, G. Multhaup, B. L. McDonald, and K. Beyreuther, "Amyloid plaque core protein in Alzheimer disease and Down syndrome.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 82, no. 12, pp. 4245–9, 1985.

[4] K. G. Mawuenyega, W. Sigurdson, V. Ovod, L. Munsell, T. Kasten, J. C. Morris, K. E. Yarasheski, and R. J. Bateman, "Decreased Clearance of CNS b -Amyloid in Alzheimer ' s Disease," vol. 330, pp. 2010, 2010.

[5] J. Hardy, "Amyloid, the presenilins and Alzheimer's disease," *Trends Neurosci.*, vol. 20, no. 4, pp. 154–159, 1997.

[6] M. P. Mattson, "Pathways Towards and Away from Alzheimer's Disease," *Nature*, vol. 430, no. 7000, pp. 631–639, 2011.

[7] O. Sofola, F. Kerr, I. Rogers, R. Killick, H. Augustin, C. Gandy, M. J. Allen, J. Hardy, S. Lovestone, and L. Partridge, "Inhibition of GSK-3 ameliorates Abeta pathology in an adult-onset Drosophila model of Alzheimer's disease.," *PLoS Genet*, vol. 6, no. 9, 2010.

[8] C. Link, "Gene expression analysis in a transgenic Caenorhabditis elegans Alzheimer's disease model," *Neurobiol. Aging*, vol. 24, no. 3, pp. 397–413, 2003.

[9] M. Citron, D. Westaway, W. Xia, and G. Carlson, "Mutant presenilins of Alzheimer's disease increase production of 42-residue amyloid β-protein in both transfected cells and transgenic mice," *Nat. Med.*, vol. 3, pp. 67–72, 1997.

[10] D. Harold, R. Abraham, P. Hollingworth, R. Sims, A. Gerrish, M. L. Hamshere, J. S. Pahwa, V. Moskvina, K. Dowzell, A. Williams, N. Jones, C. Thomas, A. Stretton, A. R. Morgan, S. Lovestone, J. Powell, P. Proitsi, M. K. Lupton, C. Brayne, D. C. Rubinsztein, M. Gill, B. Lawlor, A. Lynch, K. Morgan, K. S. Brown, P. A. Passmore, D. Craig, B. McGuinness, S. Todd, C. Holmes, D. Mann, A. D. Smith, S. Love, P. G. Kehoe, J. Hardy, S. Mead, N. Fox, M. Rossor, J. Collinge, W. Maier, F. Jessen, B. Schürmann, R. Heun, H. van den Bussche, I. Heuser, J. Kornhuber, J. Wiltfang, M. Dichgans, L. Frölich, H. Hampel, M. Hüll, D. Rujescu, A. M. Goate, J. S. K. Kauwe, C. Cruchaga, P. Nowotny, J. C. Morris, K. Mayo, K. Sleegers, K. Bettens, S. Engelborghs, P. P. De Deyn, C. Van Broeckhoven, G. Livingston, N. J. Bass, H. Gurling, A. McQuillin, R. Gwilliam, P. Deloukas, A. Al-Chalabi, C. E. Shaw, M. Tsolaki, A. B. Singleton, R. Guerreiro, T. W. Mühleisen, M. M. Nöthen, S. Moebus, K.-H. Jöckel, N. Klopp, H.-E. Wichmann, M. M. Carrasquillo, V. S. Pankratz, S. G. Younkin, P. A. Holmans, M. O'Donovan, M. J. Owen, and J. Williams, "Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease.," *Nat. Genet.*, vol. 41, no. 10, pp. 1088–93, 2009.

[11] P. Hollingworth, D. Harold, R. Sims, A. Gerrish, J.-C. Lambert, M. M. Carrasquillo, R. Abraham, M. L. Hamshere, J. S. Pahwa, V. Moskvina, K. Dowzell, N. Jones, A. Stretton, C. Thomas, A. Richards, D. Ivanov, C. Widdowson, J. Chapman, S. Lovestone, J. Powell, P. Proitsi, M. K. Lupton, C. Brayne, D. C. Rubinsztein, M. Gill, B. Lawlor, A. Lynch, K. S. Brown, P. A. Passmore, D. Craig, B. McGuinness, S. Todd, C. Holmes, D. Mann, A. D. Smith, H. Beaumont, D. Warden, G. Wilcock, S. Love, P. G. Kehoe, N. M. Hooper, E. R. L. C. Vardy, J. Hardy, S. Mead, N. C. Fox, M. Rossor, J. Collinge, W. Maier, F. Jessen, E. Rüther, B. Schürmann, R. Heun, H. Kölsch, H. van den Bussche, I. Heuser, J. Kornhuber, J. Wiltfang, M.

Dichgans, L. Frölich, H. Hampel, J. Gallacher, M. Hüll, D. Rujescu, I. Giegling, A. M. Goate, J. S. K. Kauwe, C. Cruchaga, P. Nowotny, J. C. Morris, K. Mayo, K. Sleegers, K. Bettens, S. Engelborghs, P. P. De Deyn, C. Van Broeckhoven, G. Livingston, N. J. Bass, H. Gurling, A. McQuillin, R. Gwilliam, P. Deloukas, A. Al-Chalabi, C. E. Shaw, M. Tsolaki, A. B. Singleton, R. Guerreiro, T. W. Mühleisen, M. M. Nöthen, S. Moebus, K.-H. Jöckel, N. Klopp, H.-E. Wichmann, V. S. Pankratz, S. B. Sando, J. O. Aasly, M. Barcikowska, Z. K. Wszolek, D. W. Dickson, N. R. Graff-Radford, R. C. Petersen, C. M. van Duijn, M. M. B. Breteler, M. A. Ikram, A. L. DeStefano, A. L. Fitzpatrick, O. Lopez, L. J. Launer, S. Seshadri, C. Berr, D. Campion, J. Epelbaum, J.-F. Dartigues, C. Tzourio, A. Alpérovitch, M. Lathrop, T. M. Feulner, P. Friedrich, C. Riehle, M. Krawczak, S. Schreiber, M. Mayhaus, S. Nicolhaus, S. Wagenpfeil, S. Steinberg, H. Stefansson, K. Stefansson, J. Snaedal, S. Björnsson, P. V Jonsson, V. Chouraki, B. Genier-Boley, M. Hiltunen, H. Soininen, O. Combarros, D. Zelenika, M. Delepine, M. J. Bullido, F. Pasquier, I. Mateo, A. Frank-Garcia, E. Porcellini, O. Hanon, E. Coto, V. Alvarez, P. Bosco, G. Siciliano, M. Mancuso, F. Panza, V. Solfrizzi, B. Nacmias, S. Sorbi, P. Bossù, P. Piccardi, B. Arosio, G. Annoni, D. Seripa, A. Pilotto, E. Scarpini, D. Galimberti, A. Brice, D. Hannequin, F. Licastro, L. Jones, P. A. Holmans, T. Jonsson, M. Riemenschneider, K. Morgan, S. G. Younkin, M. J. Owen, M. O'Donovan, P. Amouyel, and J. Williams, "Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease.," *Nat. Genet.*, vol. 43, no. 5, pp. 429–35, 2011.

[12] A. Association, "2014 Alzheimer's Disease Facts and Figures," *Alzheimer's Dement. J. Alzheimer's Assoc.*, vol. 10, no. 2, pp. 47–92, 2014.

[13] S. K. Van Den Eeden, "Incidence of Parkinson's Disease: Variation by Age, Gender, and Race/Ethnicity," *Am. J. Epidemiol.*, vol. 157, no. 11, pp. 1015–1022, 2003.

[14] B. Vellas, M. C. Carrillo, C. Sampaio, H. R. Brashear, E. Siemers, H. Hampel, L. S. Schneider, M. Weiner, R. Doody, Z. Khachaturian, J. Cedarbaum, M. Grundman, K. Broich, E. Giacobini, B. Dubois, R. Sperling, G. K. Wilcock, N. Fox, P. Scheltens, J. Touchon, S. Hendrix, S. Andrieu, and P. Aisen, "Designing drug trials for Alzheimer's disease: what we have learned from the release of the phase III antibody trials: a report from the EU/US/CTAD Task Force.," *Alzheimers. Dement.*, vol. 9, no. 4, pp. 438–44, 2013.

[15] I. Rogers, F. Kerr, P. Martinez, J. Hardy, S. Lovestone, and L. Partridge, "Ageing increases vulnerability to aβ42 toxicity in Drosophila.," *PLoS One*, vol. 7, no. 7, p. e40569, 2012.

[16] J. L. Chou, D. V Shenoy, N. Thomas, P. K. Choudhary, F. M. Laferla, S. R. Goodman, and G. A. M. Breen, "Early dysregulation of the mitochondrial proteome in a mouse model of Alzheimer's disease.," *J. Proteomics*, vol. 74, no. 4, pp. 466–79, 2011.

[17] P. L. Ross, Y. N. Huang, J. N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlet-Jones, F. He, A. Jacobson, and D. J. Pappin, "Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.," *Mol. Cell. Proteomics*, vol. 3, pp. 1154–1169, 2004.

[18] A. Thompson, J. Schäfer, K. Kuhn, S. Kienle, J. Schwarz, G. Schmidt, T. Neumann, R. Johnstone, A. K. A. Mohammed, and C. Hamon, "Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS.," *Anal. Chem.*, vol. 75, no. 8, pp. 1895–904, 2003.

[19] R. A. Zubarev, "The challenge of the proteome dynamic range and its implications for in-depth proteomics.," *Proteomics*, vol. 13, no. 5, pp. 723–6, 2013.

[20] P. Xu, H. Tan, D. M. Duong, Y. Yang, J. Kupsco, K. H. Moberg, H. Li, P. Jin, and J. Peng, "Stable isotope labeling with amino acids in Drosophila for quantifying proteins and modifications.," *J. Proteome Res.*, vol. 11, no. 9, pp. 4403–12, 2012.

[21] V. J. Patel, K. Thalassinos, S. E. Slade, J. B. Connolly, A. Crombie, J. C. Murrell, and J. H. Scrivens, "A comparison of labeling and label-free mass spectrometry-based proteomics approaches.," *J. Proteome Res*, vol. 8, no. 7, pp. 3752–9, 2009.

[22] K. A. Neilson, N. A. Ali, S. Muralidharan, M. Mirzaei, M. Mariani, G. Assadourian, A. Lee, S. C. van Sluyter, and P. A. Haynes, "Less label, more free: approaches in label-free quantitative mass spectrometry.," *Proteomics*, vol. 11, no. 4, pp. 535–53, 2011.

[23] Y. Lyutvinskiy, H. Yang, D. Rutishauser, and R. A. Zubarev, "In silico instrumental response correction improves precision of label-free proteomics and accuracy of proteomics-based predictive models.," *Mol. Cell. Proteomics*, vol. 12, no. 8, pp. 2324–31, 2013.

[24] J. P. C. Vissers, J. I. Langridge, and J. M. F. G. Aerts, "Analysis and quantification of diagnostic serum markers and protein signatures for Gaucher disease.," *Mol. Cell. Proteomics*, vol. 6, no. 5, pp. 755–66, 2007.

[25] T. Geiger, J. Cox, and M. Mann, "Proteomics on an Orbitrap benchtop mass spectrometer using all-ion fragmentation.," *Mol. Cell. Proteomics*, vol. 9, no. 10, pp. 2252–61, 2010.

[26] B. C. Collins, L. C. Gillet, G. Rosenberger, H. L. Röst, A. Vichalkovski, M. Gstaiger, and R. Aebersold, "Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system.," *Nat. Methods*, vol. 10, no. 12, pp. 1246–53, 2013.

[27] S. J. Geromanos, J. P. C. Vissers, J. C. Silva, C. A. Dorschel, G.-Z. Li, M. V Gorenstein, R. H. Bateman, and J. I. Langridge, "The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS.," *Proteomics*, vol. 9, no. 6, pp. 1683–95, 2009.

[28] Y. Levin, E. Hradetzky, and S. Bahn, "Quantification of proteins using data independent analysis (MS(E)) in simple and complex samples: a systematic evaluation.," *Proteomics*, pp. 3273–3287, 2011.

[29] M. D. Hoos, B. M. Richardson, M. W. Foster, A. Everhart, J. W. Thompson, M. A. Moseley, and C. A. Colton, "Longitudinal Study of Di ff erential Protein Expression in an Alzheimer ' s Mouse Model Lacking Inducible Nitric Oxide Synthase," *J. Proteome Res.*, vol. 12, pp. 4462–4477, 2013.

[30] T. Osterwalder, K. S. Yoon, B. H. White, and H. Keshishian, "A conditional tissue-specific transgene expression system using inducible GAL4.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, no. 22, pp. 12596–601, 2001.

[31] L. M. Luheshi, W. Hoyer, T. P. de Barros, I. van Dijk Härd, A.-C. Brorsson, B. Macao, C. Persson, D. C. Crowther, D. A. Lomas, S. Ståhl, C. M. Dobson, and T. Härd, "Sequestration of the Abeta peptide prevents toxicity and promotes degradation in vivo.," *PLoS Biol.*, vol. 8, no. 3, p. e1000334, 2010.

[32] A. B. Chakraborty, S. J. Berger, and J. C. Gebler, "Use of an integrated MS – multiplexed MS / MS data acquisition strategy for high-coverage peptide mapping studies," pp. 730–744, 2007.

[33] J. C. Silva, R. Denny, C. A. Dorschel, M. Gorenstein, I. J. Kass, G.-Z. Li, T. McKenna, M. J. Nold, K. Richardson, P. Young, and S. Geromanos, "Quantitative proteomic analysis by accurate mass retention time pairs.," *Anal. Chem.*, vol. 77, no. 7, pp. 2187–200, 2005.

[34] J. Grossmann, B. Roschitzki, C. Panse, C. Fortes, S. Barkow-Oesterreicher, D. Rutishauser, and R. Schlapbach, "Implementation and evaluation of relative and absolute quantification in shotgun proteomics with label-free methods.," *J. Proteomics*, vol. 73, no. 9, pp. 1740–6, 2010.

[35] J. C. Silva, M. V Gorenstein, G.-Z. Li, J. P. C. Vissers, and S. J. Geromanos, "Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition.," *Mol. Cell. Proteomics*, vol. 5, no. 1, pp. 144–56, 2006.

[36] S. Dudoit, Y. H. Yang, M. J. Callow, and T. P. Speed, "STATISTICAL METHODS FOR IDENTIFYING DIFFERENTIALLY EXPRESSED GENES IN REPLICATED cDNA MICROARRAY EXPERIMENTS," vol. 12, pp. 111–139, 2002.

[37] B. M. Bolstad, R. A. Irizarry, M. Astrand, and T. P. Speed, "A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.," *Bioinformatics*, vol. 19, no. 2, pp. 185–93, 2003.

[38] A. Universal sample preparation method for proteome analysisZougman, N. Nagaraj, and M. Mann, "Universal sample preparation method for proteome analysis," *Nat. Methods*, vol. 6, no. 5, pp. 3–7, 2009.

[39] R. P. Dator, K. W. Gaston, and P. A. Limbach, "Multiple Enzymatic Digestions and Ion Mobility Separation Improve Quanti fi cation of Bacterial Ribosomal Proteins by Data Independent Acquisition Liquid Chromatography − Mass Spectrometry," *Anal. Chem*, 2015.

[40] A. Franceschini, D. Szklarczyk, S. Frankild, M. Kuhn, M. Simonovic, A. Roth, J. Lin, P. Minguez, P. Bork, C. von Mering, and L. J. Jensen, "STRING v9.1: protein-protein interaction networks, with increased coverage and integration.," *Nucleic Acids Res.*, vol. 41, no. Database issue, pp. D808–15, 2013.

[41] J. A. Barbas, J. Galceran, I. Krah-Jentgens, J. L. de la Pompa, I. Canal, O. Pongs, and a Ferrus, "Troponin I is encoded in the haplolethal region of the Shaker gene complex of Drosophila.," *Genes Dev.*, vol. 5, no. 1, pp. 132–140, 1991.

[42] U. Nongthomba, S. Clark, M. Cummins, M. Ansari, M. Stark, and J. C. Sparrow, "Troponin I is required for myofibrillogenesis and sarcomere formation in Drosophila flight muscle.," *J. Cell Sci.*, vol. 117, no. 9, pp. 1795–805, 2004.

[43] J. E. Adams, G. S. Bodor, V. G. Davila-Roman, J. A. Delmez, F. S. Apple, J. H. Ladenson, and A. S. Jaffe, "Cardiac troponin I. A marker with high specificity for cardiac injury," *Circulation*, vol. 88, no. 1, pp. 101–106, 1993.

[44] D. Amparan, D. Avram, C. G. Thomas, M. G. Lindahl, J. Yang, G. Bajaj, and J. E. Ishmael, "Direct interaction of myosin regulatory light chain with the NMDA receptor.," *J. Neurochem.*, vol. 92, no. 2, pp. 349–61, 2005.

[45] B. DuBoff, J. Götz, and M. B. Feany, "Tau promotes neurodegeneration via DRP1 mislocalization in vivo.," *Neuron*, vol. 75, no. 4, pp. 618–32, 2012.

[46] L. M. Ittner, Y. D. Ke, F. Delerue, M. Bi, A. Gladbach, J. van Eersel, H. Wölfing, B. C. Chieng, M. J. Christie, I. A. Napier, A. Eckert, M. Staufenbiel, E. Hardeman, and J. Götz, "Dendritic function of tau mediates amyloid-beta toxicity in Alzheimer's disease mouse models.," *Cell*, vol. 142, no. 3, pp. 387–97, 2010.

[47] J. G. Sørensen and V. Loeschcke, "Decreased heat-shock resistance and down-regulation of Hsp70 expression with increasing age in adult Drosophila melanogaster", vol. 16, no. 3 pp. 379–384, 2002.

[48] T. Hoshino, N. Murao, T. Namba, M. Takehara, H. Adachi, M. Katsuno, G. Sobue, T. Matsushima, T. Suzuki, and T. Mizushima, "Suppression of Alzheimer's disease-related phenotypes by expression of heat shock protein 70 in mice.," *J. Neurosci.*, vol. 31, no. 14, pp. 5225–34, 2011.

[49] P. Mehien, C. Kretz-remy, X. Preville, and A. Arrigo, “aB-crystallin expression-mediated increase in glutathione is essential for the protective activity of these proteins against TNFa-induced cell death,” vol. 15, no. 11, pp. 2695–2706, 1996.

[50] R. T. Perry, J. S. Collins, H. Wiener, R. Acton, and R. C. P. Go, “The role of TNF and its receptors in Alzheimer ’ s disease,” vol. 22, pp. 873–883, 2001.

[51] J. Acunzo, M. Katsogiannou, and P. Rocchi, “Small heat shock proteins HSP27 (HspB1), αB-crystallin (HspB5) and HSP22 (HspB8) as regulators of cell death.,” *Int. J. Biochem. Cell Biol.*, vol. 44, no. 10, pp. 1622–31, 2012.

[52] J. M. Bruey, C. Ducasse, P. Bonniaud, L. Ravagnan, S. A. Susin, C. Diaz-Latoud, S. Gurbuxani, A. P. Arrigo, G. Kroemer, E. Solary, and C. Garrido, “Hsp27 negatively regulates cell death by interacting with cytochrome c.,” *Nat. Cell Biol.*, vol. 2, no. 9, pp. 645–52, 2000.

[53] M. E. Tóth, V. Szegedi, E. Varga, G. Juhász, J. Horváth, E. Borbély, B. Csibrány, R. Alföldi, N. Lénárt, B. Penke, and M. Sántha, “Overexpression of Hsp27 ameliorates symptoms of Alzheimer’s disease in APP/PS1 mice.,” *Cell Stress Chaperones*, vol. 18, no. 6, pp. 759–71, 2013.

[54] H.-D. Wang, P. Kazemi-Esfarjani, and S. Benzer, “Multiple-stress analysis for isolation of Drosophila longevity genes.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 34, pp. 12610–5, 2004.

[55] P.-C. Liao, H.-Y. Lin, C.-H. Yuh, L.-K. Yu, and H.-D. Wang, “The effect of neuronal expression of heat shock proteins 26 and 27 on lifespan, neurodegeneration, and apoptosis in Drosophila.,” *Biochem. Biophys. Res. Commun.*, vol. 376, no. 4, pp. 637–41, 2008.

[56] S. R. Pfeffer, “Structural clues to Rab GTPase functional diversity.,” *J. Biol. Chem.*, vol. 280, no. 16, pp. 15485–8, 2005.

[57] M. Geppert and T. C. Südhof, “RAB3 and synaptotagmin: the yin and yang of synaptic membrane fusion.,” *Annu. Rev. Neurosci.*, vol. 21, pp. 75–95, 1998.

[58] J. Somsel Rodman and A. Wandinger-Ness, “Rab GTPases coordinate endocytosis.,” *J. Cell Sci.*, vol. 113 Pt 2, pp. 183–92, 2000.

[59] G. von Mollard, B. Stahl, and C. Li, “Rab proteins in regulated exocytosis,” *Trends Biochem. Sci.*, vol. 19, no. 4, pp. 164–168, 1994.

[60] C.-C. Chan, S. Scoggin, D. Wang, S. Cherry, T. Dembo, B. Greenberg, E. J. Jin, C. Kuey, A. Lopez, S. Q. Mehta, T. J. Perkins, M. Brankatschk, A. Rothenfluh, M. Buszczak, and P. R. Hiesinger, “Systematic discovery of Rab GTPases with synaptic functions in Drosophila.,” *Curr. Biol.*, vol. 21, no. 20, pp. 1704–15, 2011.

[61] P. H. Reddy, G. Mani, B. S. Park, J. Jacques, G. Murdoch, W. Whetsell, J. Kaye, and M. Manczak, “Differential loss of synaptic proteins in Alzheimer ’ s disease : Implications for synaptic dysfunction,” vol. 7, pp. 103–117, 2005.

[62] E. R. Graf, R. W. Daniels, R. W. Burgess, T. L. Schwarz, and A. DiAntonio, “Rab3 dynamically controls protein composition at active zones.,” *Neuron*, vol. 64, no. 5, pp. 663–77, 2009.

[63] O. M. Schlüter, F. Schmitz, R. Jahn, C. Rosenmund, and T. C. Südhof, “A complete genetic analysis of neuronal Rab3 function.,” *J. Neurosci.*, vol. 24, no. 29, pp. 6629–37, 2004.

[64] A. R. English and G. K. Voeltz, “Rab10 GTPase regulates ER dynamics and morphology.,” *Nat. Cell Biol.*, vol. 15, no. 2, pp. 169–78, 2013.

[65] E. E. Kelly, F. Giordano, C. P. Horgan, F. Jollivet, G. Raposo, and M. W. McCaffrey, "Rab30 is required for the morphological integrity of the Golgi apparatus.," *Biol. Cell*, vol. 104, no. 2, pp. 84–101, 2012.

[66] J. Zhou, D. R. Jones, D. M. Duong, A. I. Levey, J. J. Lah, and J. Peng, "Proteomic analysis of postsynaptic density in Alzheimer's disease.," *Clin. Chim. Acta.*, vol. 420, pp. 62–8, 2013.

[67] D. Mastroeni, L. Chouliaras, A. Grover, W. S. Liang, K. Hauns, J. Rogers, and P. D. Coleman, "Reduced RAN expression and disrupted transport between cytoplasm and nucleus; a key event in Alzheimer's disease pathophysiology.," *PLoS One*, vol. 8, no. 1, p. e53349, 2013.

[68] Y. Ren, W. X. Hong, F. Davey, M. Taylor, J. Aiton, P. Coote, F. Fang, J. Yao, D. Chen, J. X. Chen, D. Y. Shi, and F. J. Gunn-Moore, "Endophilin I expression is increased in the brains of Alzheimer disease patients," *J. Biol. Chem.*, vol. 283, no. 9, pp. 5685–5691, 2008.

[69] T. Maier, M. Güell, and L. Serrano, "Correlation of mRNA and protein in complex biological samples," *FEBS Lett.*, vol. 583, no. 24, pp. 3966–3973, 2009.

[70] J. L. Mott, S. Kobayashi, S. F. Bronk, and G. J. Gores, "mir-29 regulates Mcl-1 protein expression and apoptosis.," *Oncogene*, vol. 26, no. 42, pp. 6133–6140, 2007.

[71] M. R. Fabian, N. Sonenberg, and W. Filipowicz, "Regulation of mRNA translation and stability by microRNAs.," *Annu. Rev. Biochem.*, vol. 79, pp. 351–379, 2010.

[72] M.-S. Kim, S. M. Pinto, D. Getnet, R. S. Nirujogi, S. S. Manda, R. Chaerkady, A. K. Madugundu, D. S. Kelkar, R. Isserlin, S. Jain, J. K. Thomas, B. Muthusamy, P. Leal-Rojas, P. Kumar, N. A. Sahasrabuddhe, L. Balakrishnan, J. Advani, B. George, S. Renuse, L. D. N. Selvan, A. H. Patil, V. Nanjappa, A. Radhakrishnan, S. Prasad, T. Subbannayya, R. Raju, M. Kumar, S. K. Sreenivasamurthy, A. Marimuthu, G. J. Sathe, S. Chavan, K. K. Datta, Y. Subbannayya, A. Sahu, S. D. Yelamanchi, S. Jayaram, P. Rajagopalan, J. Sharma, K. R. Murthy, N. Syed, R. Goel, A. A. Khan, S. Ahmad, G. Dey, K. Mudgal, A. Chatterjee, T.-C. Huang, J. Zhong, X. Wu, P. G. Shaw, D. Freed, M. S. Zahari, K. K. Mukherjee, S. Shankar, A. Mahadevan, H. Lam, C. J. Mitchell, S. K. Shankar, P. Satishchandra, J. T. Schroeder, R. Sirdeshmukh, A. Maitra, S. D. Leach, C. G. Drake, M. K. Halushka, T. S. K. Prasad, R. H. Hruban, C. L. Kerr, G. D. Bader, C. A. Iacobuzio-Donahue, H. Gowda, and A. Pandey, "A draft map of the human proteome.," *Nature*, vol. 509, no. 7502, pp. 575–81, 2014.

[73] M. Wilhelm, J. Schlegl, H. Hahne, A. Moghaddas Gholami, M. Lieberenz, M. M. Savitski, E. Ziegler, L. Butzmann, S. Gessulat, H. Marx, T. Mathieson, S. Lemeer, K. Schnatbaum, U. Reimer, H. Wenschuh, M. Mollenhauer, J. Slotta-Huspenina, J.-H. Boese, M. Bantscheff, A. Gerstmair, F. Faerber, and B. Kuster, "Mass-spectrometry-based draft of the human proteome.," *Nature*, vol. 509, no. 7502, pp. 582–7, 2014.

[74] P. Verstreken, O. Kjaerulff, T. E. Lloyd, R. Atkinson, Y. Zhou, I. A. Meinertzhagen, and H. J. Bellen, "Endophilin mutations block clathrin-mediated endocytosis but not neurotransmitter release," *Cell*, vol. 109, no. 1, pp. 101–112, 2002.

Longitudinal Label-Free Quantitative Proteomics of Chronic Aβ Induction and Wild-Type Ageing in *D. melanogaster* to Deconstruct Ageing as a Major Risk Factor for Alzheimer's Disease

3.0 Abstract

Alzheimer's disease (AD) is a complex neurodegenerative process whose aetiology is thought to be increased levels of Aβ peptide species. The biggest risk factor for late-onset AD is age, though the cellular basis behind this link is thought to be multifaceted and requires further investigation. Using a well-studied, inducible, *D. melanogaster* AD model, longitudinal profiling of brain specific protein expression levels, in response to wild-type ageing and chronic Aβ peptide induction, was performed. Protein abundance levels for approximately 2000 proteins were consistently quantified across successive time points using a label-free data independent acquisition LC-MS/MS approach, termed $HDMS^E$. Around 15% of the quantified brain proteome across the six time points measured was found to be significantly altered due to ageing. A large number of metabolic proteins were found to be differentially expressed suggesting an alteration of mitochondrial function during ageing. Protein-protein interaction data further complements this theory with multiple proteins involved in the electron transport chain identified as being downregulated upon increasing age. Quantification of the brain proteome across the complete *Drosophila* lifespan, in the presence of chronic Aβ peptide, found little direct overlap with our ageing control when considering statistically significant changes in abundance. Soft cluster analysis of protein profiles, however, indicated similar biological processes to be affected between the two conditions. Cross-referencing our data with mRNA profiles from a previous study successfully highlighted common expression changes at the mRNA and protein level. These data represent a high confidence set of cellular responses to Aβ toxicity in the presence of increasing age for follow up functional studies. Direct comparison of abundance levels measured in the two conditions identified a smaller number of differentially regulated proteins that can be confidently attributed as a direct cellular response to Aβ peptide toxicity. Based on these identifications gene ontology analysis indicates alterations in chaperone levels, neurogenesis and learning/memory. Together our data provide a unique resource to better dissect the complex aetiology of AD and the functional links with age-related cellular alterations.

3.1 Introduction

Alzheimer's disease has a well-known but largely uncharacterised preclinical phase where AD-related neuropathological abnormalities are present but cognitive ability remains normal [1], [2]. This preclinical phase is best described by an accumulation of APP-derived Aβ peptide species within the brain, although there is still debate as to whether this is the main aetiology of the disease or part of a more complex pathological process. Indeed a number of high profile late-stage clinical trials targeting the production or aggregation of amyloid-beta peptide have failed to reduce symptoms [3]–[5]. A major problem for such drug development is the gradual nature of the disease pathology with extracellular deposition of Aβ aggregates known to predate clinical symptoms by at least ten years. Limitations in obtaining representative samples during the preclinical phase have prevented the characterisation of early stage AD beyond what can be determined from cerebrospinal fluid (CSF) or plasma [6], [7]. The analysis of CSF or plasma provides a readily available and relatively uninvasive method for developing biomarkers to improve the ability to diagnose AD but is limited in the information it can provide in terms of disease progression. With evidence now indicating that the oligomers that precede the characteristic extracellular plaques to be the more toxic Aβ species in AD, it is important that more is done to dissect the cellular consequences of early stage Aβ accumulation and to correlate this with the later stages of the disease [8], [9]. Model organisms such as *Drosophila melanogaster* and *Mus musculus* are used to study human diseases and are particularly common in the study of neurodegenerative diseases and ageing [10], [11]. Although vertebrae models have pathological hallmarks that more closely reflect human disease disadvantages include longer experimental timescales and cost-intensive handling. Shorter lived animal model systems including *Drosophila* provide the ability to perform relatively short lifespan and negative geotaxis experiments that can facilitate large-scale screening experiments. A downside is that such models can lack some pathophysiological characteristics of AD. Despite being relatively short lived, *Drosophila* have a well-studied, complex brain and central nervous system which along with the ease of genetic manipulation make them an ideal model for neurodegeneration. Indeed, there are a range of fly models for neurodegenerative diseases including Parkinson's disease, amyotrophic lateral sclerosis and AD [12]–[14]. For AD there are several different models that over-express either Aβ peptide species or tau protein [15]. Recently, an adult-onset fly model for AD, which uses an inducible pan-neuronal

driver, elav GeneSwitch, to over-express the Arctic mutant Aβ42 peptide in the fly central nervous system, has been developed and used to highlight that the inhibition of GSK-3 ameliorates Aβ pathology [16]. Expression of Aβ42 within fly neurons causes protein aggregation into an insoluble, fibrillary form similar to those seen in human AD patients. Additionally, an increase in mortality is observed alongside neuronal dysfunction, as measured by negative geotaxis (climbing), without an observed loss of neuronal cells. It was also used to identify age-related vulnerability to Aβ42 induction in aged flies, successfully validating the model for use in studies aiming to better understand age-related susceptibility to AD [17].

Mass spectrometry-based quantitative proteomics is the chosen method for characterising cellular disease states in an unbiased manner [18]–[20]. Proteomics has been used extensively in studies of neurodegenerative diseases to identify differentially expressed proteins/modifications by comparing disease to an appropriately matched control state [21]–[23]. Results either provide information to better model the pathology of the disease or can be used to identify potential biomarker candidates for improved disease diagnosis. Label-free quantification currently offers the technical precision and accuracy required to study biological systems (see Chapter 1.14) [24].

Time and sample handling constraints have limited the majority of discovery proteomic experiments to focus on a single time point that is thought to best represent the disease in question. This is an obvious simplification but allows for follow up targeted experiments across a more representative time period for the proteins identified as being of biological interest. The continued development of mass spectrometry instrumentation has reduced sample analysis time whilst concurrently increasing proteome coverage [25], [26]. It is now feasible to undertake a longitudinal analysis of disease progression by mass spectrometry, where, analysis timescales are measureable by weeks rather than the months older instruments would require to achieve comparable results.

Here quantitative time-course analysis of an adult-onset *D. melanogaster* model for Alzheimer's disease was performed. Differentially regulated proteins were identified across the fly lifespan in response to the neuronal specific expression of Arctic Aβ42 peptide. *Drosophila* brains were analysed using a recently developed label-free mass spectrometry technique HDMS^E (see Chapter 1.16). Using our DIA approach approximately 2000 proteins were successfully quantified across successive time points and in both conditions. The data represent a unique resource for dissecting brain specific

proteome alterations to ageing and Aβ peptide toxicity. Soft clustering of protein time-course profiles in response to normal ageing identified a number of biological processes that had a coordinated proteome response. Alterations in processes involved in energy production, sexual reproduction and oxidoreductase activity were found to be enriched in specific clusters. Indeed, around 15% of all quantified proteins were significantly altered ($q < 0.05$) across the time-course. Protein-protein interaction analysis identified a complex network of differential regulation. In comparison, although clustering of protein profiles in response to chronic Aβ expression suggested similar biological processes to be altered, little overlap of differentially regulated proteins between the two datasets was observed. Regulated proteins were involved in a wide range of cellular process including neuronal stem cell maintenance, oxidoreductase activity and the protein folding/translation. Our results were successfully validated by comparison with mRNA levels measured in a similar *Drosophila* AD model. A subset of proteins where both mRNA and protein levels are significantly altered by Aβ peptide misfolding was successfully highlighted. Finally direct comparison of expression levels measured in the two conditions identified a small number of proteins that are altered as a direct response to Aβ expression. Functional analysis identified proteins implicated in chaperoning, neurogenesis and learning all of which are thought play key roles in neurodegeneration.

Overall our unbiased and longitudinal approach has successfully collected data that represents a unique resource for follow-up functional studies on both normal ageing and Aβ peptide toxicity. Biological processes previously implicated in neurodegeneration have been highlighted and better characterised by our longitudinal approach. As mass spectrometric instrumentation improves further sill, in terms of speed and sensitivity, time-course characterisation of protein expression levels will become commonplace.

3.2 MATERIALS AND METHODS

3.2.1 Fly stocks and Maintenance

All fly stocks were maintained at 25 °C on a 12:12 hrs light:dark cycle at constant humidity on a standard sugar-yeast (SY) medium (15 g/l agar, 50 g/l sugar, 100 g/l autolysed yeast, 100 g/l nipagin and 3 ml/l propionic acid). Adult onset neuronal-specific expression of Arctic mutant Aβ42 peptide was obtained with the use of the elav GeneSwitch (elavGS)-UAS system.[27] This line was derived from the original elavGS 301.2 line and was a gift from Dr H.Tricoire. The UAS-ArcAβ42 line was obtained as a generous gift from Dr D Crowther. All transgene lines were backcrossed six times into the w118 genetic background. For all experiments female flies carrying UAS-ArcAβ42 were crossed with male flies expressing elavGS. Expression of Aβ peptide in female progeny was achieved by treatment with mifepristone (RU486) added to the SY medium.

3.2.2 Ageing-span and Lifespan

Flies were raised at a standard density on SY medium in 200 mL bottles. Two days post eclosion, to ensure flies were once-mated, female flies were split to experimental vials containing SY medium with or without RU486. For the lifespan, flies were split at a density of 10 flies per vial whilst for the ageing experiments flies were split at a density of 25 flies per vial. Flies were transferred to fresh food three times a week at which point deaths were scored for the lifespan. Lifespan data are presented as cumulative survival curves.

3.2.3 Protein Isolation and Tryptic Digestion

Brains were obtained for induced and non-induced flies by dissection at each chosen time point (six for non-induced and four for induced). Time points were chosen to span the majority of the lifespan. Dissections were carried out following CO_2 anesthetisation on ice cold PBS. For each biological repeat 10 brains were collected, snap frozen and stored at -80 °C until required.

All chemicals unless otherwise stated were obtained from Sigma Aldrich, MO, USA. Proteins from brain samples were extracted by homogenisation on ice into 50 µl of 50 mM Ammonium Bicarbonate 0.25% Rapigest (Waters Corporation, MA, USA) and 10 mM DTT. Proteins were solubilised and disulphide bonds reduced by heating at 80 °C for 20 minutes. Free cysteine residues were alkylated by the addition of IAA to a final concentration of 20 mM and incubation at room temperature in the dark for 20 minutes. A 1 µl aliquot was taken from each sample for protein concentration determination using the Qubit protein assay kit (Thermo Fisher, MA, USA). Samples were diluted with 50 mM Ammonium bicarbonate to a final concentration of 0.1% Rapigest. Tryptic digestion was carried out overnight at 37 °C overnight by the addition of sequencing grade trypsin at a 1:50 enzyme:protein ratio. To ensure efficient digestion a further addition of trypsin (Promega, Southampton, UK) at a ratio of 1:100 was performed the following morning and each sample incubated for a further hour. Detergent was removed post-digestion by sample acidification with 0.1% formic acid and incubation at 60 °C for 60 minutes. Insoluble detergent and sample debris was removed by centrifugation at 14,000 x *g* for 30 minutes. Supernatant was collected, lyophilised to dryness and stored at -80 °C.

3.2.4 LC-HDMS^E Analysis

Nanoscale UPLC separation of tryptic peptides was performed using a nanoAcquity UPLC system (Waters Corporation, MA, USA) equipped with a UPLC HSS T3 1.7 µm, 75 µm x 250 mm analytical reverse phase column (Waters Corporation, MA, USA). Prior to peptide separation, 300 ng of tryptic peptides were loaded onto a 2G, V/V 5 µm, 180 µm x 20 mm reverse phase trapping column (Waters Corporation, MA, USA) at 5 µl/min for 3 minutes. Mass spectrometric analysis of tryptic digests was performed using a Synapt G2S-Si HDMS mass spectrometer (Waters, Manchester, UK) equipped with a T-Wave-IMS device. Mass measurements were made in positive ion-mode ESI with the instrument operated in resolution mode with a typical resolving power of 20,000 FWHM (full width at half maximum). Prior to analysis the time-of-flight mass analyser was externally calibrated with the fragment ion of [Glu1]-Fibrinopeptide B (GFP) from *m/z* 175 to 1285. The data were post-acquisition lock mass corrected using the double charged monoisotopic ion of GFP. To achieve lock mass correction a 100 fmol/ul solution of GFP was infused at a 90° angle to the analytical sprayer. This

reference sprayer was sampled every 60 seconds. Accurate mass LC-IM-MS data were collected in the DIA mode of analysis, $HDMS^E$. IMS was performed by applying a constant wave height of 40 V whilst a constant wave velocity of 650 m/s was maintained. Wave heights within the trap and transfer were both set at 4 V whilst the wave velocities were 311 and 175 m/s respectively. MS data were acquired over 50-2000 *m/z for* each mode. Spectral acquisition time for each mode was 0.5 s with a 0.015 interscan delay. During the low energy MS mode data were acquired whilst applying a constant collision energy of 4 eV within the transfer travelling wave region. High energy MS/MS data were acquired by ramping the collision energy within the transfer region, post IMS, between 15 V and 40 V. One cycle of low and elevated energy data were acquired every 1.1 s. To ensure that ions with *m/z* less than 350 observed in the LC-MS data were derived exclusively from peptide fragmentation within the transfer region the radio frequency applied to the quadrupole mass analyser was adjusted to optimise transmission within the region of 350 – 2000 Da. Each biological replicate was analysed three times. Samples were kindly run by Lee Gethings and Chris Hughes at Waters, Manchester.

3.2.5 Digestion Reproducibility

Proteins were extracted and digested from *Drosophila* brain tissue as described earlier in the experimental section. Prior to digestion, lysate concentration was adjusted to 1 µg/µl and three proteins, alcohol dehydrogenase, luciferase and lysozyme were spiked in at concentrations of 300 fmol/µl, 167 fmol/µl and 958 fmol/µl respectively. For each of the three proteins two surrogate isotopically labelled peptides were spiked into the lysate at as close to equimolar concentration as possible. The sample was aliquoted and digested in triplicate overnight. Post-digestion samples were acidified with formic acid and dried to completion. Prior to analysis samples were resolublised at 0.5 µg/µl in 0.1% formic acid. Samples were analysed in technical triplicates using the same conditions as detailed in Chapter 2 with the only difference being that a total of 1 µg was loaded on column.

Data analysis was carried out based on extracted ion chromatograms of unlabelled and labelled peptides. Peak top intensities were compared after combining scans across the entire chromatographic peak. Prior to peak top intensity calculation data were smoothed (3 scan window, x2 smooth), baseline subtracted (Polynomial order:10, 80% below curve, 0.10 Tolerance) and centroided (90% top peak). Processing steps for all samples were identical.

3.2.6 Time-course Data Analysis

Data were processed using Progenesis QI v2.0 (Waters Corporation, MA, USA). Data were imported into Progenesis to generate a 3D representation of the data (*m/z*, RT and peak intensity). Samples were then time aligned with the software allowed to automatically determine the best reference run from the dataset. Briefly alignment is based upon the placement of landmarks called alignment vectors where each vector connects the location of a peptide ion on the reference run to the location of the same ion on the run to be aligned. Using a non-linear algorithm the retention times between the reference run and the run to be aligned are mapped. Initially a gross alignment is calculated by finding the smallest number of retention time shifts that bring the two chromatograms into close alignment. Following this, small changes to the alignment are made with the impact of the adjustments measured in order to achieve the best alignment possible. Peak picking was performed on MS level data. A peak picking sensitivity of 4 (out of 5) was set. To ensure consistent peak picking across multiple runs and to reduce the number of missing values across replicates and conditions, an aggregate peak data set is created from aligned runs. This aggregate data set includes RT, *m/z* and drift time information allowing the utilisation of the increased peak capacity offered by $HDMS^E$ analysis. The process is somewhat analogous to 'match between runs' which is implemented in the MaxQuant software developed by J. Cox and colleagues [28].

Post alignment and peak picking, peptide features were noise subtracted and ion current integrated by the Apex3D algorithm. The pep3D algorithm subsequently de-isotoped and charge state reduced features into resultant EMRTs. Peptide features were tentatively aligned with their respective fragment ions based primarily on the similarity of their chromatographic and mobility profiles. Requirements for features to be included in post-processing database searching were as follows; 300 counts for low energy ions, 50 counts for high energy ions and 750 counts for precursor EMRT integrated intensity.

Subsequently, data were searched against the Uniprot canonical *Drosophila* database (20,049 sequences) appended with common contaminants. Trypsin was specified as the enzyme of choice and a maximum of two missed cleavages were permitted. Carbamidomethyl (C) was set as a fixed modification whilst oxidation (M) and acetylation (N-term) were set as variable modifications. Requirements for successful peptide identification were at least 1 fragment ion whilst for protein

identification a minimum of 3 fragment ions and 1 peptide was required. Protein level FDR was set to a 4% threshold based upon the search of an automatically generated reversed database.

Three different methods of relative protein quantification were tested prior to any data analysis. Hi3, all peptide summation and all non-conflicting (unique) peptide summation were compared. Prior to quantification proteins were grouped according to peptide identifications. Where more than one protein was identified with common peptides, the protein with fewer peptides was subsumed into that with the greater number of peptides and not reported for quantification. Where proteins could not be distinguished based on detected peptides a protein group was reported. To control for instrumental sensitivity run to run variability and LC loading variation, peptide intensities were normalised prior to quantification. Briefly, based upon the assumption that the majority of peptides do not change in abundance across different biological samples and conditions, a scalar factor is calculated. To calculate the scalar factor a reference sample is first calculated. Following this, peptide intensity ratios (sample/ratio) are calculated for each sample and log10 transformed to ensure a normal distribution. The mean and median of this distribution of ratios is used to define the data suitable for calculating the scalar factor. This ensures that outlying ratio values do not skew the data. Calculated scalar factors are applied to each sample on a non-logarithmic scale.

3.2.7 Statistical Testing

Differential expression across individual time-series was tested for using an ANOVA and the optimal discovery procedure. ANOVA analysis was performed in Progenesis QI where time point was set as the variable. The optimal discovery procedure was implemented within the Bioconductor package, Extraction of Differential Gene Expression (EDGE) [29]. For each time-course dataset the degrees of freedom were set to one less than the number of time points sampled. For all tests, FDR was controlled by the calculation of q values from the p value output. Briefly, q values control the FDR in large datasets, where multiple hypothesis tests are performed, based on the observed p value distribution. By plotting a histogram of all calculated p values and calculating the height of the histogram, at the point at which the larger p values become evenly distributed, an FDR adjustment factor can be calculated. Differential expression was defined as $q \leq 0.05$. All quantitative data and statistical results are provided in the appendix file.

Comparison of the two time-course datasets (Aβ vs age matched control) was achieved using a two-sided student's t-test with unequal variance. Expression values were compared between the two conditions at matched time points. Differential expression was defined as a protein having a p value < 0.05 in 2 or more time points out of the 4 tested.

3.2.8 Enrichment Analysis

Enrichment analysis was performed in both DAVID and Panther where biological process, molecular function, cellular compartment and panther class (Panther) were tested for enrichment.[30], [31] Briefly, although using slightly different algorithms, the two software, calculate enrichment based on the number of gene ontology terms present in your chosen background (usually organism specific). A p value can be calculated by comparing the number proteins with a specific gene ontology term present in your data to the number present in the background, in relation to the total number of proteins tested. Enrichment analysis within DAVID was performed using all identified proteins as a background and enrichment was defined as a p value < 0.005 or an FDR < 0.05. Panther analysis was performed against the *D. melanogaster* background, as defined in Panther, and enrichment was defined as a p value < 0.05 as bonferonni correction.

3.2.9 Cluster Analysis

Prior to grouping expression profiles, proteins were filtered to only include those that were found to have a high level of expression change across time points. Briefly, the correlation between protein intensity and quantitative CV was calculated using the same approach as detailed in Chapter 2. Effect sizes for quantified proteins were calculated from the largest fold change observed, again as detailed in Chapter 2. Proteins with an effect size larger than 0.7 were taken forward for cluster analysis using the R package Mfuzz [32]. The default fuzzification factor of 1.25 was used in all cases. Cluster numbers were optimised by varying the cluster number from 5-15 and plotting the resultant principle component analysis (PCA) plot. The minimum number of clusters that best described the data variation was taken forward for analysis. Results clustering analysis are provided in the appendix file.

3.2.10 Protein-Protein Interaction Analysis

Proteins of interest were queried in STRING for known protein-protein interactions [33]. All types of interaction were included. Interactions were filtered based on confidence of the interaction with either medium confidence (score of ≥ 0.4) or high confidence (score of ≥ 0.7) taken forward for analysis. Score are calculated by the STRING database. Data from all sources (including experimental, computational prediction and text collections) is weighted and integrated prior to confidence calculations. Data were imported into Cytoscape for visualisation [34]. Where cluster analysis was performed the EAGLE algorithm was used to define clusters and was implemented within the ClusterViz application available through Cytoscape [35].

3.3 Results & Discussion

3.3.1 Comparison of GeLCMS^E and In-Solution-HDMS^E Protocols

Previously *Drosophila* brain tissue was analysed with the use of SDS-PAGE to aid protein solubilisation and to increase protein identification numbers by sample fractionation prior to LC-MS^E analysis on a Synapt HDMS instrument (Chapter 2). Following the release of the Synapt G2 by Waters Ltd all subsequent Synapt instruments have the ability to separate peptides by ion mobility, post LC separation and prior to ToF mass analysis, with a data independent acquisition approach, commonly referred to as HDMS^E. The Synapt G2S a few years later saw the introduction of the StepWave technology, an off-axis ion source which increases signal by up to 25-fold. The StepWave consists of a stacked-ring ion guide that is designed with two stages that are set off-axis from each other. This offset design ensures the removal of neutral molecules that enter the source from the atmosphere thereby allowing a far larger source orifice to be used without the issue of neutral contaminant build up. With the increase in peak capacity afforded by HDMS^E and the increased sensitivity provided by the introduction of the StepWave technology it was expected that analysis of brain tissue by HDMS^E could provide a similar number of protein identifications than the GeLCMS^E protocol with an order of magnitude reduction in analysis time.

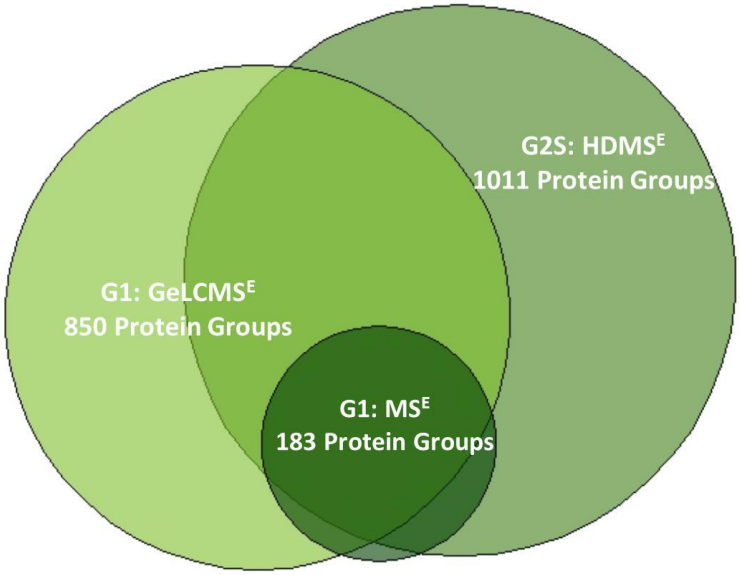


Figure 3.1 Comparison of protein identifications from three MS protocols used to analyse *Drosophila* brain tissue

Two samples that were representative of the samples to be analysed during the planned time-course experiment were prepared from w118 *Drosophila* of mixed ages. To test the HDMS^E protocol a total of 10 brains were dissected and pooled prior to sample homogenisation, protein extraction and tryptic digestion. To test the GeLCMS^E protocol a total of 10 brains were pooled, homogenised in x 2 sample buffer and separated on an SDS-PAGE gel as detailed in Chapter 2. Fractionated and un-fractionated samples were separated by identical C18 columns and across the same 90 minute analytical RP-LC gradient to reduce possible sources of bias.

Results demonstrated that HDMS^E analysis of *Drosophila* brain tissue identified over 100 more proteins than the GeLCMS^E protocol in a total analysis time of 120 minutes compared to 1200 minutes (after filtering for proteins identified in all three technical replicates), **Figure 3.1**. Interestingly, when all identified proteins from technical replicates were considered (not filtering based on replication) the HDMS^E and GeLCMS^E protocol identified a more similar number of proteins at 5% FDR. The data, therefore, provides evidence for the shotgun approach providing more reproducible protein identifications, which is beneficially for the statistical analysis of label-free quantitative results. The results demonstrate the applicability of HDMS^E to provide adequate proteome coverage at an acquisition time short enough to make the analysis of time-course data feasible. Without prior SDS-

PAGE fractionation, analysis by MS^E alone was only able to identify around 200 proteins in all three technical replicates. Surprisingly when the quantitative reproducibilities of technical replicates were compared, the two methods gave very similar results with an average CV of 20%. It would be expected that the gel-based method would give less precise results due to the added source of experimental error during gel fractionation. The result of this would however be most pronounced when measuring biological reproducibility which was not examined in this comparison.

Another result of note from this experiment was that the $HDMS^E$ and $GeLCMS^E$ protocols identified a large number of proteins unique to the respective protocols. It can be seen in, **Figure 3.1**, that this is not observed when comparing the two shotgun methods (G1 vs G2S-i), indicating that the differences are largely caused by sample preparation rather than during analysis. Differences in the amount of detergent used in the two protocols (4% vs 0.25%) are likely to cause the solubilisation of proteins (and subsequent efficient digestion) with different hydrophobicities. Additionally carrying out the digestion in-gel in contrast to in-solution is likely to cause difference in digestion efficiencies and peptide recovery. Alternatively some differences could be caused by the fact that the two protocols carry out the fractionation at different stages of the sample preparation protocol. Whilst gel-based methods fractionate based on protein size, $HDMS^E$ fractionates based on peptide size and charge. This could be expected to effect the peptides/proteins identified and indeed in Chapter 2, it was observed that the $GeLCMS^E$ method over represents larger proteins, which was hypothesised to be caused by the gel-based separation employed.

3.2.2 Assessment of Digestion Reproducibility

A common assumption in proteomic sample preparation protocols is that the tryptic digestion of a sample has proceeded to completion, with all proteins being enzymatically digested into peptides at equimolar concentration to the original protein concentration. Another assumption in quantitative proteomics is that the digestion is reproducible and that there is therefore little to no error associated with digestion variability. This second assumption is paramount to the accuracy of label-free relative quantification and has been previously shown to be an over simplification [36]. In-gel-based protocols, digestion efficiency benefits from the ability to use high concentrations of detergent in the presence of a protease. In-solution protocols are limited to concentrations of detergent/denaturant that retain

enzyme activity and are therefore more susceptible to being incomplete or irreproducible. We wanted to assess the suitability of an in-solution digest protocol commonly employed within the lab and to confirm that reproducibility across biological samples was high enough for use in a label-free quantitative study. Our approach used isotopically labelled peptides as internal standards for proteins spiked into a *Drosophila* brain tissue background at known concentrations. Three proteins, hen egg white lysozyme, yeast alcohol dehydrogenase (ADH) and firefly luciferase, along with two isotopically labelled peptides per protein were employed as detailed in **Table 3.1**.

Table 3.1 Synthetic peptides used to assess the extent and reproducibility of tryptic digestion. Three standard proteins were spiked into a *Drosophila* brain homogenate background. Peptide position, in reference to its parent protein, and details of isotopic labelling are included

Protein	Peptide Position	Sequence and isotopic labelling
ADH	T25	EALD[13C9,15N-F]FAR
ADH	T28	V[13C5,15N-V]GLSTLPE[13C6,15N-I]YEK
LUCIFERASE	T36	VVP[13C9,15N-F]FEAK
LUCIFERASE	T47	E[13C6,15N-I]VDYV[13C3,15N-A]SQVTTAK
LYSOZYME	T5	HG[13C6,15N-L]DNYR
LYSOZYME	T7	FESN[13C9,15N-F]NTQATNR

Each protein was spiked into the background sample at different concentrations to test the applicability of the digestion protocol over a concentration range equating to roughly an order of magnitude. ADH was spiked in at a final concentration of 300 fmol/µl, luciferase at 167 fmol/µl and lysozyme at 958 fmol/µl. The background sample was prepared at a concentration of 1 µg/µl (total protein) as the *in-house* digestion protocol is commonly performed at a final concentration between 0.1-1 µg/µl. Three aliquots were prepared from a single stock for tryptic digestion. The digestion protocol was based upon the use of the mass spectrometry compatible surfactant, Rapigest and post-digestion two isotopically labelled peptides per protein were spiked into samples at an equimolar concentration to their respective protein analogues. Isotopic standards had been quantified by amino acid analysis to ensure the correct concentration assignment. Each sample was analysed in triplicate to facilitate the calculation on technical precision. After data acquisition, extracted ion chromatograms (XICs) were used to calculate peptide peak height intensity for ratio calculations. Continuum peptide

peaks were smoothed, subjected to baseline subtraction and were centroided. An example of the raw data and processing steps is shown in **Figure 3.2**. Alongside ratio measurements an assessment of signal to noise (*S/N*) was carried out for each peptide pair based upon the monoisotopic peak of the isotopically labelled peptide analogue, **Table 3.2.**

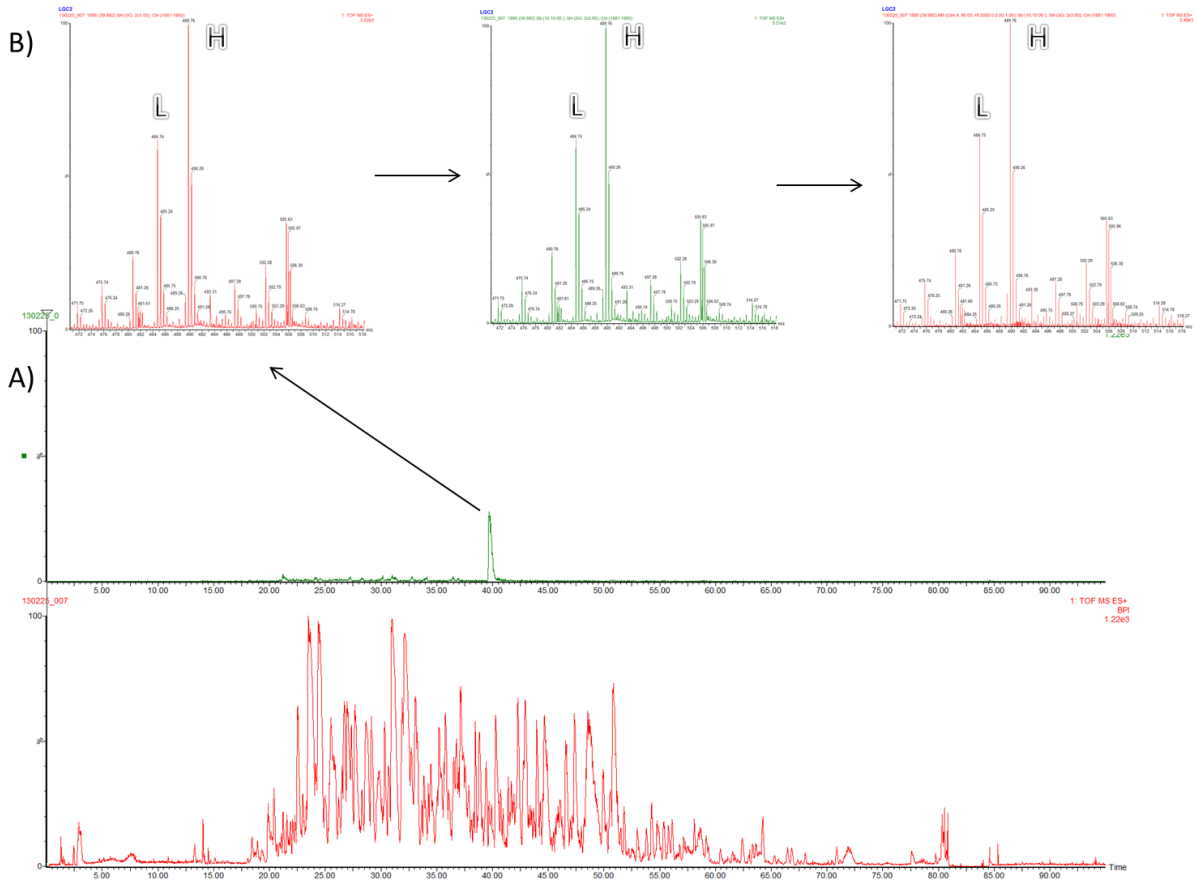


Figure 3.14 Data analysis steps for the calculation of digestion efficiency and reproducibility. **A)** An extracted chromatogram for the spiked peptide is combined to form the average spectra for light and heavy peptides. **B)** Data is smoothed, background subtracted and centroided. Ratios are then calculated from peak intensities

Signal to noise was calculated within Masslynx. All spiked peptides and their unlabelled counterparts were successfully identified and quantified with signal to noise varying from 2 up to 80. Interestingly the signal to noise did not appear to correlate with peptide concentration and was dependent on either the extent of enzymatic release and/or peptide ionisation efficiency. This is not unexpected as these are two properties that are commonly considered when choosing peptides for targeted protein quantification by SRM or parallel reaction monitoring (PRM). Reproducibility of digestion was found to be high for all peptides included in the study with an average coefficient of variance of only 5%. If the

average standard error of the mass spectrometry ratio measurement is considered (4%), then the digestion clearly contributes a small percentage of uncertainty for relative quantitative proteomics by MS.

Table 3.2 Summary of results from digestion reproducibility assessment

Peptide Position	S/N	Average Ratio	STDEV	CV (%)
T25	80	0.63	0.012	1.9
T28	34	1	0.05	5.2
T36	2	0.55	0.05	9.3
T47	32	0.79	0.03	4
T5	3	0.72	0.006	0.9
T7	38	0.7	0.06	9.7

The biggest source of error was peptide to peptide variability with a ratio range of 0.45. Although this should not impact relative quantification these results should be considered if the data were required to measure absolute amounts of protein. Absolute quantification is dependent on complete or close to complete sample digestion for accurate results. As an example, if protein quantification was based on the peptides producing low ratios (T36, T25) alone then the data would underestimate protein concentration for the respective parent proteins.

3.2.3 Experimental Design

As previously mentioned, Alzheimer's disease is characterised by its long gestation period and its strong link with ageing as a risk factor [37]. Studies to better understand the disease should therefore ideally be able to probe the development of the disease across an organisms lifetime and link this to possible age-related changes. This is an obvious limitation of AD studies on human patients. The use of an inducible system such as the elavGS-UAS construct, as used in this study, allows an experimental design such that it was possible to interrogate longitudinal protein expression levels of both normal ageing and chronic Aβ42 expression at multiple time points. This facilitated the comparison of two conditions to detect expression changes that were attributable to Aβ peptide toxicity, age-related physiological changes or were due to a combination of the two.

The Aβ flies used in this experiment carry one copy of the transgene expressing the Arctic mutant of human Aβ42. This mutation is associated with decreased levels of Aβ40 & Aβ42 in patient plasma and has been shown to form protofibrils at a much higher rate than wild-type Aβ [38], [39]. Similarly in *Drosophila* this mutation has been shown to cause the peptide to almost exclusively aggregate and accumulate into an insoluble fibril form [40]. Due to the toxic nature of the Arctic mutation, the RU486 induced flies (Aβ) have a median survival far shorter than the non-RU486 induced flies (control). The first four time points at which samples were collected were chosen to span the Aβ fly lifespan. Subsequently results from this experiment were representative of the cumulative effects of chronic Aβ42 expression, **Figure 3.3A**. A further two time points that covered the remaining lifespan of the control flies were included. These were most representative of normal age-related decline. Indeed the final two time points capture the point at which mortality first begins to increase (time point 5) and the time at which mortality is close to its most pronounced (time point 6). If the lifespan of the Aβ flies is considered it can be observed that time points 3 & 4 capture a similar level of mortality to time points 5 & 6 for the control flies.

AD is a neurodegenerative disease and we were therefore interested in profiling brain-specific protein expression. Analysis was performed on dissected *Drosophila* brain tissue, an approach not commonly used when studying *Drosophila* models [41], [42]. The benefit of this protocol was two-fold; firstly it ensured that measurements were truly representative of neurological changes occurring, whilst secondly, it reduced sample complexity through the removal of proteins specifically expressed in the eye, cuticle and fat body. For each time point four biological repeats were included and each of these were then analysed three times to ensure statistical confidence. Dissections for each time point were conducted within the same hour each day to reduce circadian effects on protein levels. Following protein extraction and tryptic digestion, resultant peptide solutions were separated based on hydrophobicity by RP-UPLC, **Figure 3.3C**, prior to $HDMS^E$ analysis on a Synapt G2S-i mass spectrometer. $HDMS^E$ is a DIA method that separates peptides by ion mobility, post LC separation and prior to MS analysis, **Figure 3.3D**. The increase in system peak capacity afforded by IM has been shown to increase peptide identification rates by up to 50% in comparison to MS^E analysis.[43] All data were processed using Progenesis v2.0 including non-linear retention time alignment of runs and peptide level intensity normalisation, **Figure 3.3E**. Data were processed in three separate cohorts to allow for different questions to be probed within each dataset using the same settings each time. The

three different data cohorts were all data collected from Aβ flies, all data collected from control flies and finally the data as an entirety. Prior to label-free quantification proteins were grouped based on the parsimony principle, a form of Occam's razor, where protein identification are reported as the minimum identifications that can be used to account for all peptides detected. Progenesis offer three main modes of label-free quantification, Hi3, all peptides and all unique peptides. The all peptides quantification method utilises all peptides matched to a protein, post protein grouping, whilst Hi3 uses the top three most ionising peptides per protein. We were interested in comparing the quantitative precision of the three methods to determine which would provide the highest level of precision. It can be seen from **Figure 3.4A**, that although Hi3 and all unique peptides provide a similar level of precision, the use of all peptides for relative quantification clearly provides higher reproducibility with over 80% of quantified proteins having a biological CV of 50% or less. The increased precision is most likely a combined effect of increased protein intensity signal for low abundant proteins and the averaging out of peptide intensity fluctuations for proteins with a large number of identified peptides. The all peptides method, although offering higher precision, is likely to offer less accurate results due to the use of shared peptides for quantification. It was noted that a large number of quantified proteins had no unique peptides assigned, with quantification solely based on peptides that are shared between multiple proteins (shared peptides). The effect of including shared peptides in quantification calculations is shown in **Figure 3.4B**. Here the all peptide method of quantification was found to quantify over a 1000 more proteins than the two alternative methods. As the unique peptide method and all peptide method quantify using an inherently similar method, the proteins quantified only by the all peptide method represent proteins for which there are no unique peptides detected. As we were more interested in quantification accuracy across biological conditions than precision we opted to quantify using the Hi3 method. Analysis found subsequent quantitative data to have a high level of technical and biological linearity at >0.92 R^2 and >0.89 R^2 respectively, **Figure 3.3F**.

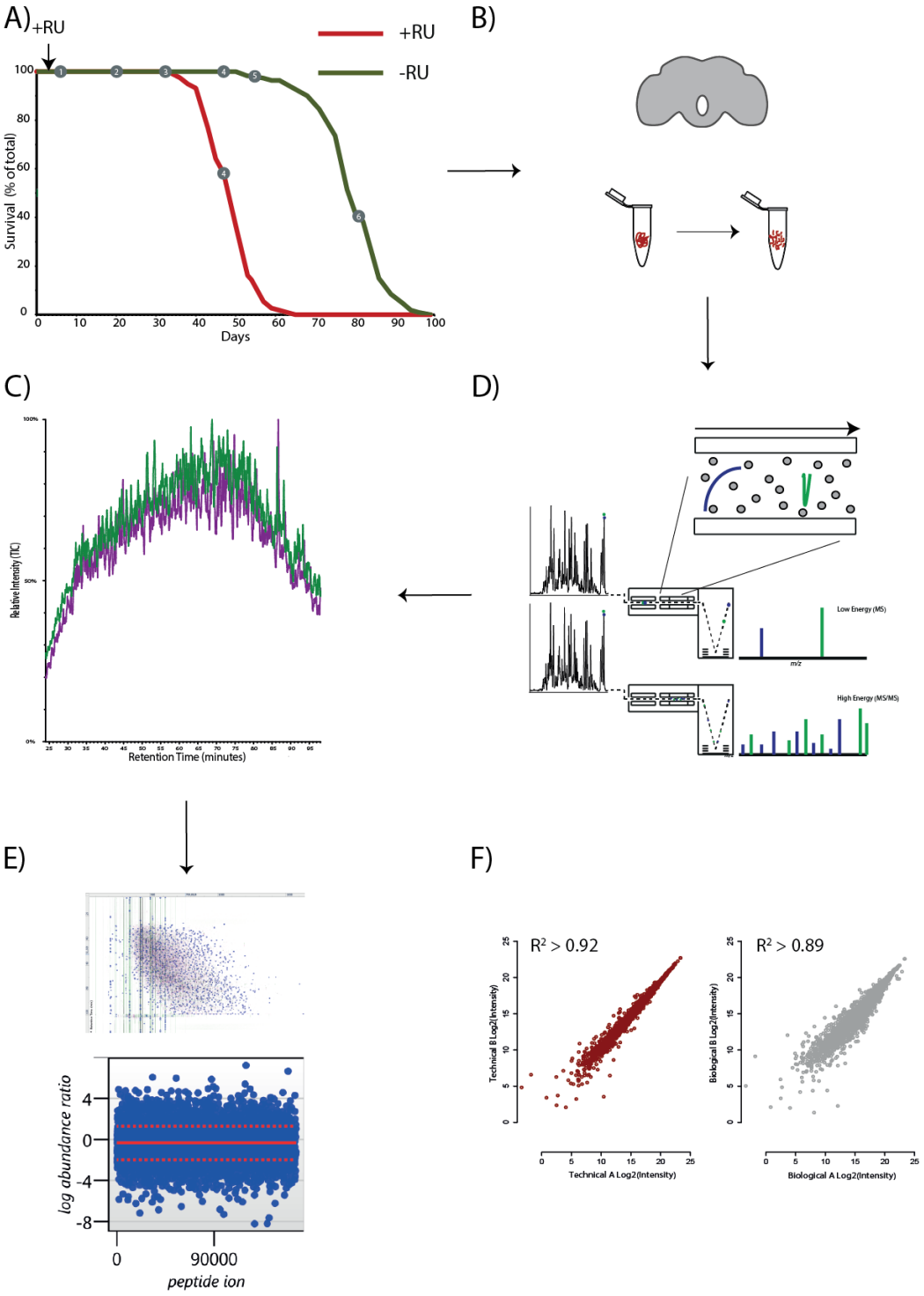


Figure 3.3 Workflow for the longitudinal quantitative proteomic analysis of Aβ induced toxicity in comparison to wild-type ageing. **A&B)** Brain tissue was dissected at 4 (Aβ induced) and 6 (ageing control) time points that spanned > 80% of the two conditions lifespan. **C&D)** Post-protein extraction and digestion peptide solutions were separated by UPLC prior to HD-MS^E analysis. **E)** Raw data were processed in Progenesis which included retention time alignment, and peptide intensity normalisation and protein inference. **F)** Normalised data were found to have a high level of technical and biological reproducibility

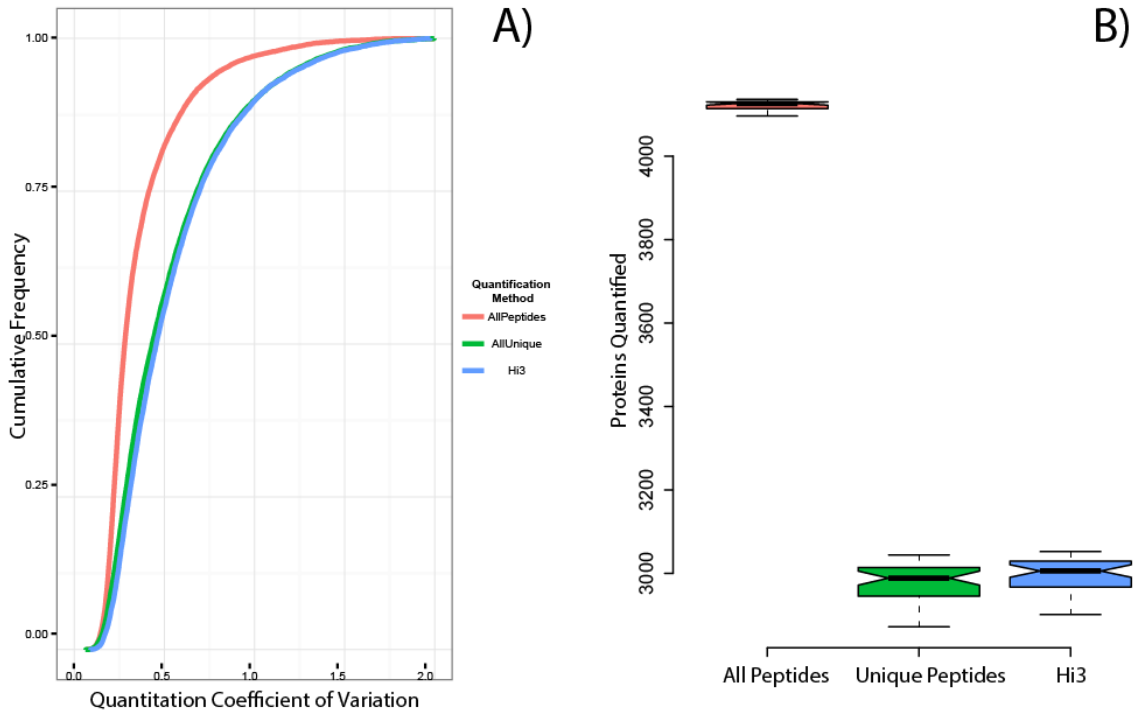


Figure 3.4 Comparison of three methods for relative label-free quantification. A) Cumulative frequency plot of protein intensity CV across repeat measurements (technical and biological). B) Number of proteins successfully quantified

2.3.4 Peak Picking Optimisation

Apex3D, the algorithm within PLGS that identifies features within the low/high energy data and subsequently clusters precursors with their respective fragment ions, requires three thresholds for defining what constitutes a true peptide feature. A high energy and low energy peak height intensity threshold must be met for a peak to be defined and an integrated area under the peak threshold also needs to be reached. Although suggested settings are provided, the best thresholds can be sample dependent and should be optimised prior to analysis. Somewhat counterintuitively reducing the thresholds does not always increase the number of proteins that are identified from the downstream database search. This is most likely due to the false discovery rate threshold being reached more quickly as a result of incorrectly defining noise as peptide and fragment features. Choosing the correct values for optimal protein identifications based on the raw data is not possible as the input values do

not readily relate to raw peak intensities. Apex3D thresholds instead represent post-processing values and can only be optimised by repeat data processing whilst varying processing parameters. Using the freely available software PLGS Threshold Inspector, high energy (HE) and low energy (LE) thresholds were optimised for maximum protein identifications, **Figure 3.5A**.

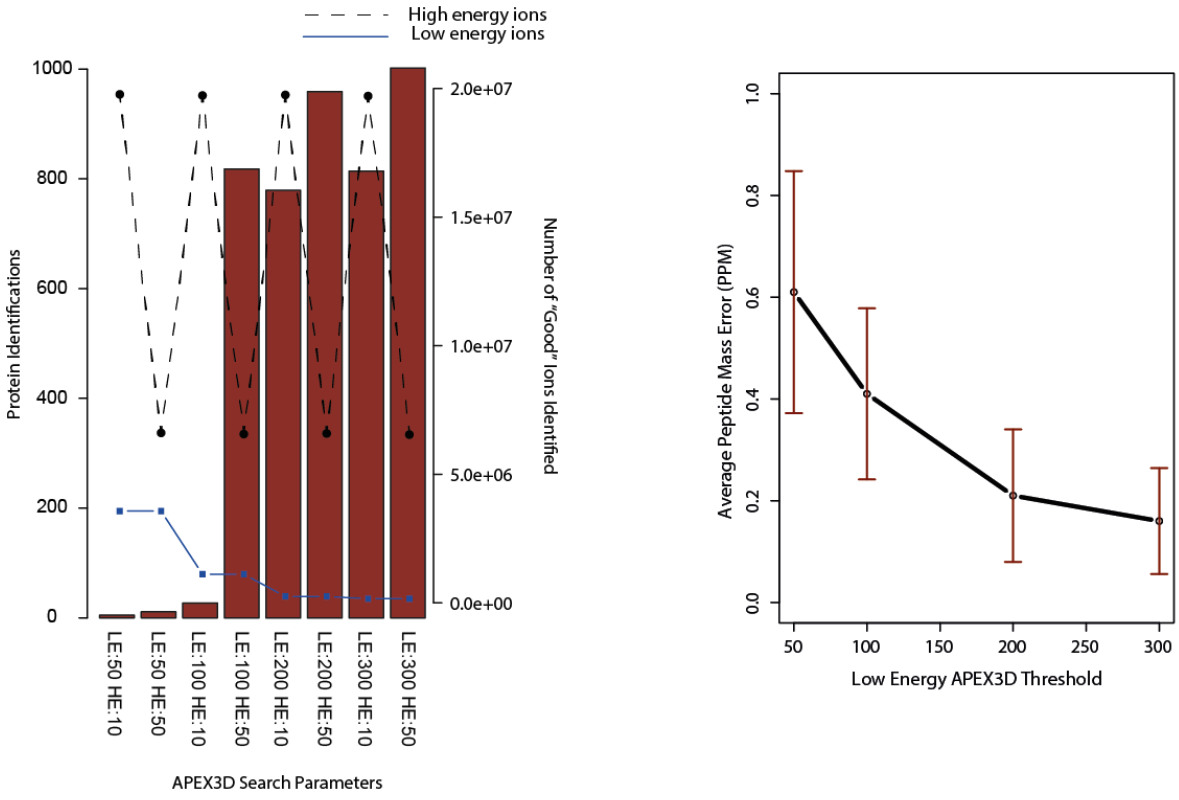


Figure 3.5 Optimisation of Apex3D peak picking settings for maximum protein identifications. **A)** Protein identification rates at four different low energy thresholds and two different high energy thresholds. **B)** Peptide mass error as a function of low energy Apex3D thresholds

It was found for the very low thresholds, LE = 50 and HE = 10 minimal protein identifications were made. This number rapidly increased when using the thresholds of LE=100 and HE=50 suggesting that these values represent a rough *S/N* threshold for the data. This threshold is validated by **Figure 3.5B** where it can be seen that increasing the LE threshold from 50 to 100 resulted in the sharpest increase in peptide mass accuracy. An increase in peptide mass accuracy is indicative of defined features being true peptides rather than isobaric noise.

2.3.5 Experimental Features

A number of features were tested to ensure that the data obtained was of sufficient quality to be taken forward for statistical and biological analyses. These included number of missed cleavages, precursor charge, ppm error and chromatographic peak width, **Figure 3.6**. The efficiency of proteolytic digestion has been previously shown to be an important factor in proteome coverage and quantitative precision [44]. One measure of the extent of tryptic digestion is to compare the proportion of semi-tryptic peptides that are identified from the database search. As seen in **Figure 3.6A** over 70% of peptides identified from the Aβ-induced time-course were fully tryptic peptides, a proportion that compares well with previous literature [45]. An alternative method for assessing the extent of digestion, thus providing a second degree of confidence, is the observed charge state distribution for identified peptides. An efficient tryptic digestion should produce peptides that predominately retain 2 or 3 positive charges during ESI. Data shown in **Figure 3.6B** clearly follows this trend with the vast majority of identified peptides being either doubly or triply charged. An alternative parameter that can influence quantitative precision is chromatographic separation. Label-free quantification is calculated

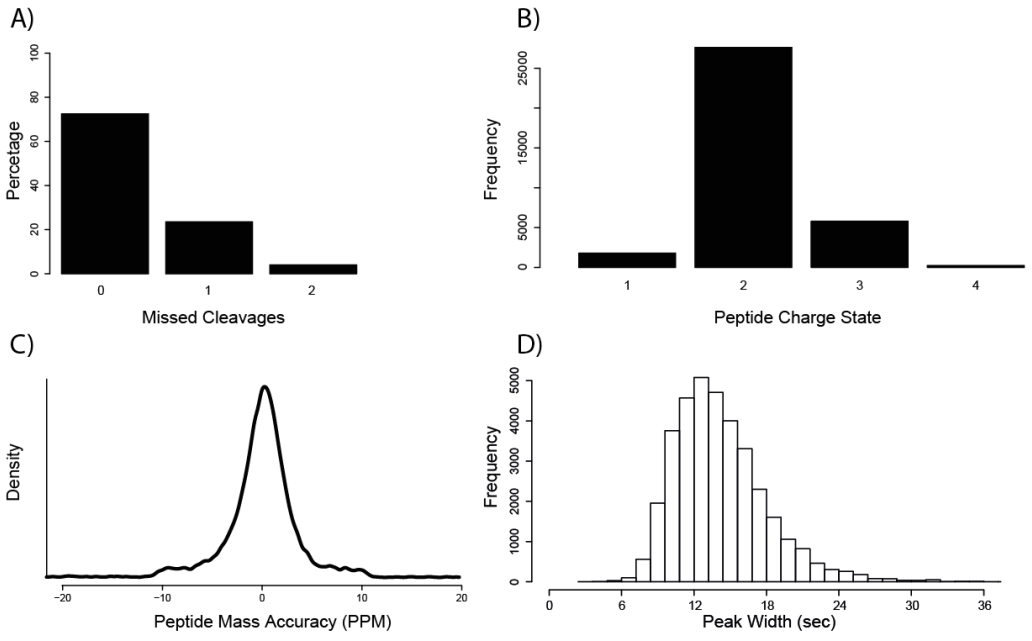


Figure 3.6 QC metrics taken from the Aβ time-course. **A)** Peptide missed cleavage distribution. **B)** Peptide charge state distribution. **C)** Peptide mass accuracy density plot. **D)** Peptide (Pass one only) peak width (FWHM) histogram.

using the area under the chromatographic peak and is therefore strongly dependent on chromatography quality. UPLC separation of peptide solutions has been previously shown to produce

peaks which have a width (FWHM) of between 6-18 seconds [46]. Our data compares favourably to this, **Figure 3.6D**, with the vast majority of peptides eluting with a peak width of around 12 second and little to no peptides having a peak width > 30 seconds. Potential sources of experimental error were examined with the use of a scatterplot matrix, whereby all protein intensities across biological samples are correlated, **Figure 3.7**. Prior analysis demonstrated technical reproducibility to be high ($R^2 > 0.92$) and were therefore not included in the plot for clarity. Reproducibility across five orders of magnitude was seen to be high: however it was observed that the best correlated samples were not biological repeats. High quantitative correlation was found between samples processed (sample homogenisation and tryptic digestion) on the same day indicating that the highest source of error within the analysis workflow was day to day sample preparation variability. Within **Figure 3.7**, two examples of this high correlation between samples processed on the same day are highlighted (indicated by grey and blue shading). These data is contrary to previous results assessing digestion reproducibility within the same day. The data therefore highlights the importance of assessing and potentially optimising day to day sample preparation variation in a label-free workflow. Experimental design was such that the influence of this variability on experimental bias had been minimised. Biological repeats were processed on different days, i.e. processing all time points from one biological repeat together. This ensured that the day-to-day variability only increased measurement error as opposed to the biasing of quantitative comparisons.

2.3.6 Assessment of Progenesis Peak Picking Algorithm

Label-free quantification is complicated by run-to-run LC peptide separation variability. Although LC reproducibility is now far more robust than in the past, peptide elution times can vary by seconds to minutes, with peptide-to-peptide variability not necessarily being a linear relationship. Retention time alignment is commonly performed across a dataset to improve the reliability and reproducibility of label-free results. A number of different algorithms have been developed to align peptide retention times across analytical runs. Two such algorithms have been developed with $HDMS^E$ and MS^E as their main applications. The first, developed by the Tenzer lab, implements retention time alignment within the IsoQuant software and aims to improve the reproducibility of quantitative results derived

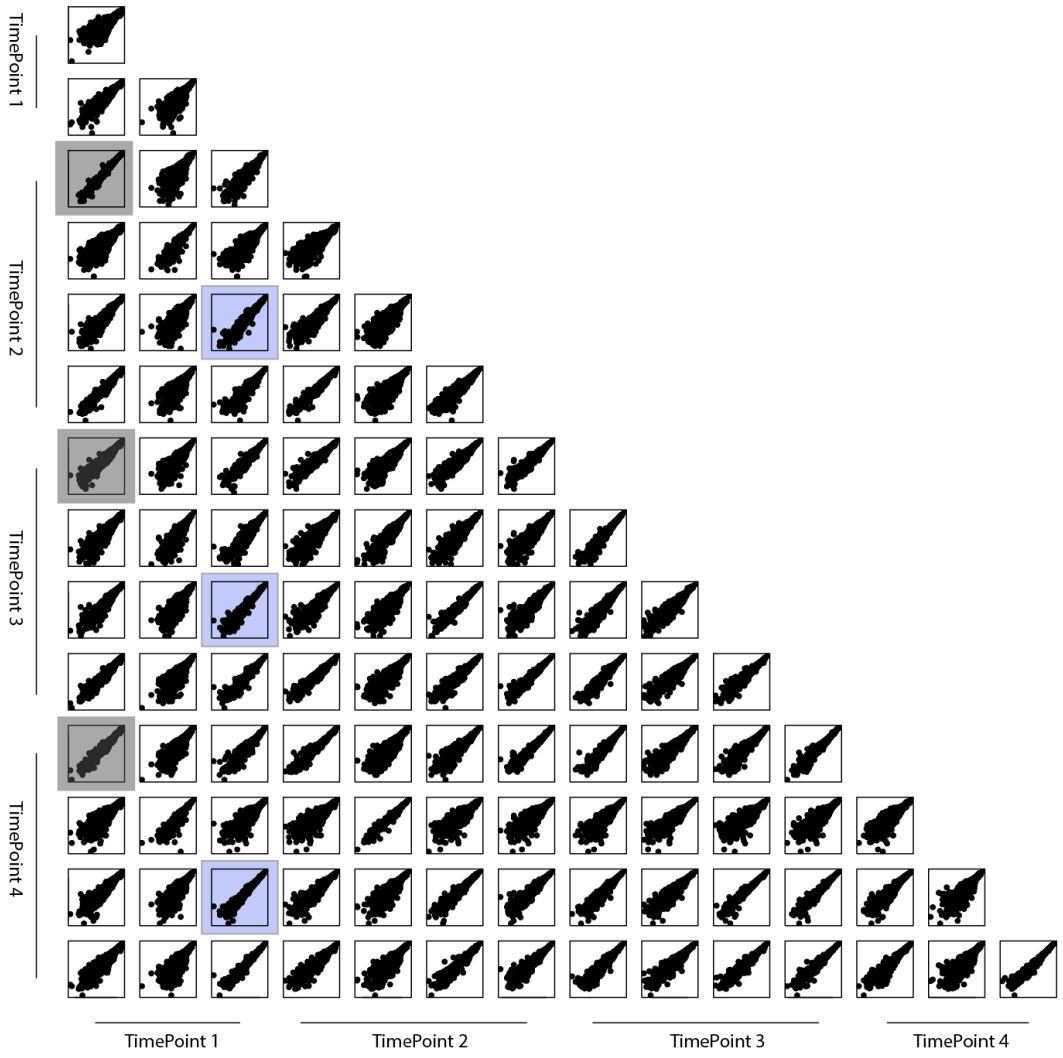


Figure 3.7 Log_2 protein intensity matrix scatterplot of Aβ+ time-course data. Each scatterplot represents a biological sample (protein intensity averaged from technical replicates). Highlighted in grey and blue are two examples of different biological conditions, prepared (homogenisation and digestion) on the same day displaying high correlation.

from MS^E derived data [46]. Here a Dynamic-Time-Warping algorithm aligns all runs to a reference run that is chosen based upon the highest number of detected signals. The second algorithm is built into the Synapter program developed by the Lilley lab which facilitates the merging of MS^E and $HDMS^E$ quantitative data [47]. Synapter performs alignment by LOESS fitting, to model retention time deviations, when comparing two samples. Retention time deviations are then corrected according to this model in a similar approach to those used for peptide intensity normalisation. Progenesis uses a third method to improve peptide LC time alignment. A 2D representation of each MS run is created by plotting *m/z* against RT. Alignment is based upon the placement of alignment vectors which act as landmarks that connect a peptide ion location from the two runs being aligned. By placing a number of

vectors within the 2D data representation a non-linear map can be produced. Alignment is then performed in two stages, firstly a small number of retention time shifts that reduce the vector deviation to an acceptable amount are calculated and then this initial alignment is optimized by performing stepwise adjustments and assessing the quality of alignment until an optimal condition is identified.

After alignment Progenesis performs an additional peak picking step prior to invoking the algorithm APEX3D which is also used in PLGS. In order to minimise the problem of missing data, which is common in proteomic datasets, an aggregate dataset is created from all aligned runs. This data is then applied to each sample to transfer protein identifications in a similar manner to the 'match between runs' application employed within the MaxQuant software [48]. A benefit of using Progenesis with $HDMS^E$ data is that the aggregate dataset also employs ion mobility drift time as an extra data dimensionality resulting in increased peak matching.

The benefits of alignment and peak picking within Progenesis in comparison with use of PLGS on its own were examined, **Figure 3.8**. Initially the number of proteins identified at a 4% protein level FDR across a set of triplicate analytical replicates were compared. Analysis by PLGS was found to consistently identify over 1000 proteins in each replicate however when the data were filtered based on replication (3 out of 3) this number reduced to little over 900. The use of the Progenesis peak picking algorithm reduced the number of missing values to only 1 across the triplicate analysis, demonstrating the power of identification transfer in proteomics datasets. A notable difference in the use of PLGS alone to the use of Progenesis is that unlike the PLGS APEX3D algorithm, peak picking within Progenesis estimates baseline noise to improve peak identification rates from the raw data. A result of this is that unlike APEX3D there are no intensity thresholds for peptide identification within Progenesis. Indeed it was noted that some peptides identified within Progenesis were well below the 750 count threshold set within PLGS. The effects of this peak processing difference was examined by filtering Progenesis data for proteins that had an integrated intensity of ≥ 750 counts multiplied by the number of peptides identified. Interestingly it was found that less than 10 protein identifications were found to be below this threshold indicating that the main data metric influencing the presence of missing values was incomplete or low quality fragmentation data. Finally the effects of identification transfer across the dataset as a whole were measured. To do this protein identifications rates for the triplicate technical replicates were compared but after peak picking was performed, using aggregate

data produced from the entire time course. It can be seen that when using data from the time-course as an entirety protein identification numbers increase by over 3-fold. The data were filtered as previously described, based upon total protein intensity, resulting in an 8% reduction in protein identifications (312 proteins). This increase in low intensity protein identifications, in comparison to previous results where protein identifications were transferred from the same biological sample, most likely represents proteins that demonstrate a degree of expression change across the time-course experiment.

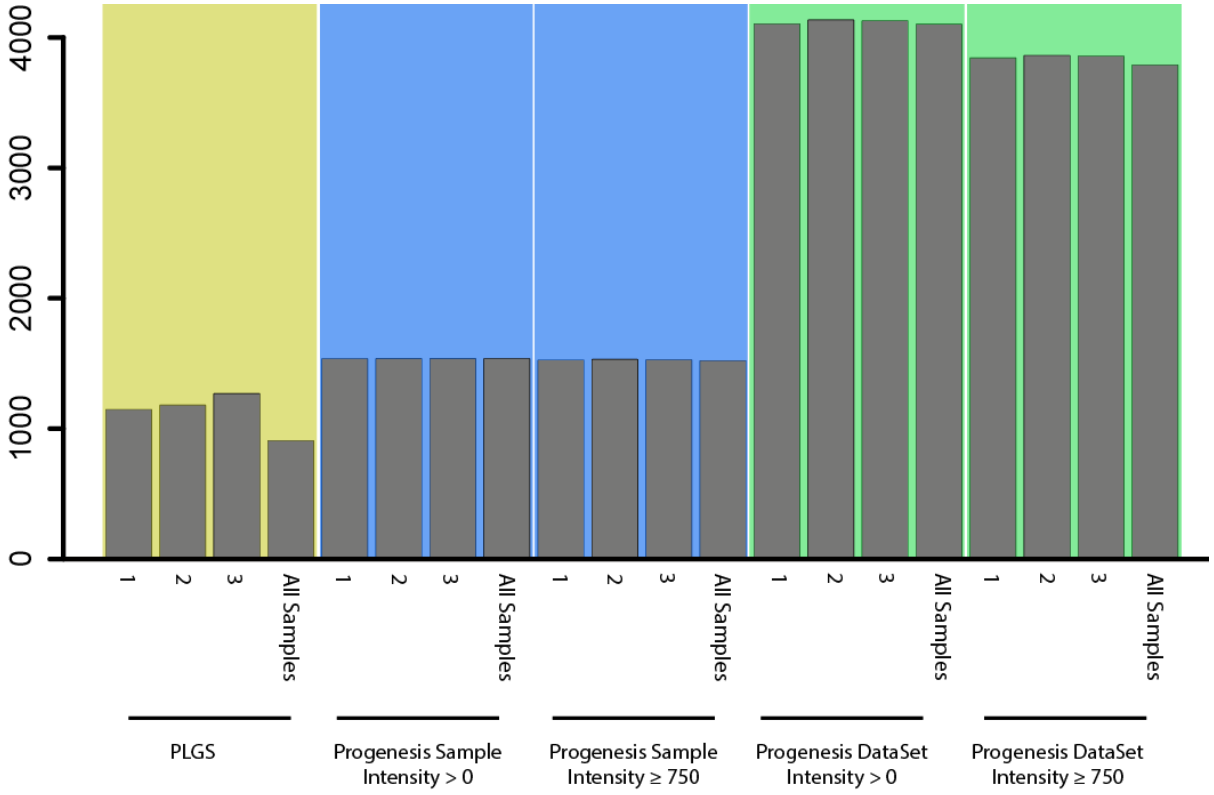


Figure 3.8 Comparison of protein identifications for three technical replicates, analysed by PLGS alone or by Progenesis

3.3.7 In-Solution Digest Provides Protein Identifications That are Representative of Starting Material

Previous mass spectrometric analysis of *Drosophila* brain tissue using a GeLCMS^E workflow was shown to be representative of starting material. Comparison of the physicochemical properties of proteins identified by GeLCMS^E to those genes detected by microarray analysis found little difference (see Chapter 2, **Figure 3.2**). In this study the use of an in-solution digestion protocol, although providing more reproducible sample preparation than a gel-based method, did not facilitate the use of high concentrations of detergent for protein solublisation. The same analysis as performed in Chapter 2 was conducted to compare the results from the in-solution protocol , the GeLCMS^E protocol and the micrroarray analysis, **Figure 3.9**. Interestingly it was observed that the use of lower concentrations of detergent (0.2% vs 4%) did not drastically reduce the representation of membrane proteins. For two other properties tested, pI and biological process gene ontology, minimal differences were observed between all three methods. Prior analysis had identified the biggest difference between the GeLCMS^E and microarray data to be the molecular weight distribution of proteins/genes identified. It was hypothesised that this was due to the use of gel seperation pior to MS analysis providing a higher resolving power for low molecular weight proteins than for larger proteins. The data derived from the in-solution protocol provides the ideal basis for testing this hypothesis. If indeed the gel-based protein separation did lead to the observed bias towards lower molecular weight proteins this should not be observed within the in-solution data set. **Figure 3.7D** confirms that as expected, the in-solution protocol did identify more high molecular weight proteins in comparison to the GeLCMS^E, however the protocol still underepresented proteins with a molecular weight > 25000 Da in comparison with the transcriptomics data. The results suggest that the apparent bias towards lower molecular weight proteins for proteomic analysis may in fact be sample-dependent rather than a result of instrumental/sample preparation bias. The results may be representative of a true difference between the expression levels of mRNA and the subsquent protein translation rates. However, without the comprehensive detection of proteins expressed within *Drosophila* brain tissue this cannot be confirmed.

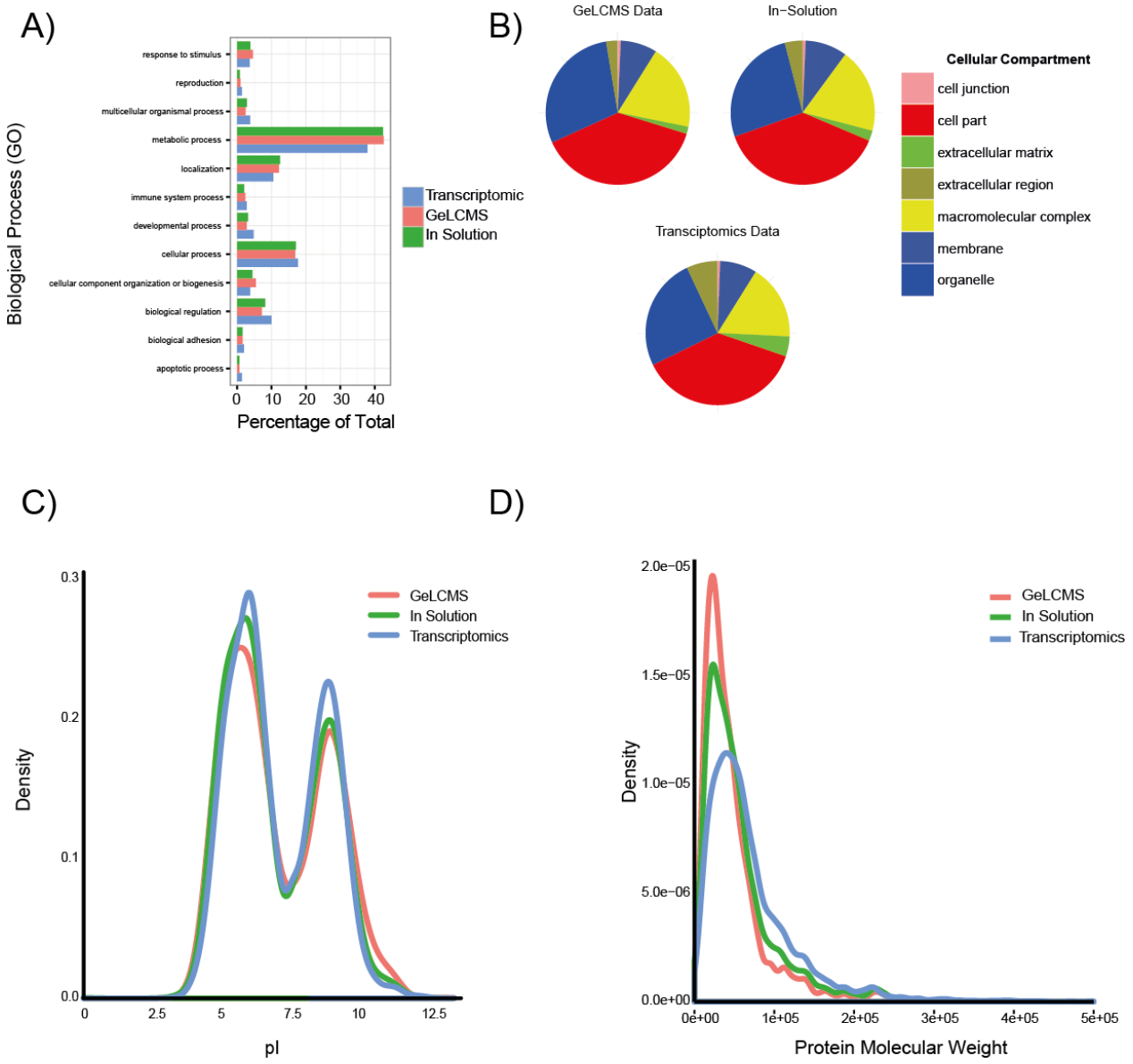


Figure 3.9 Comparison of protein gene ontology classifications and physiochemical properties of proteins/genes identified by either GeLCMS[E], HDMS[E] or microarray analysis of *Drosophila* brain tissue. **A & B)** Proteins classified by biological process or cellular compartment respectively. **C)** Density plot of protein pI distribution. **D)** Density plot of protein molecular weight distribution

3.2.8 Peptide Filtering Post Database Search

The PLGS software uses a protein-centeric database search that has been shown to identfiy up to 20% more identfications than other peptide-centric database search algorithms [49]. A result of this is that PLGS controls the FDR at the protein level but not at the peptide level. As protein identfication rates have increased in proteomics it has now become commonplace to control both the protein and peptide level FDR. In order to control the peptide level FDR, peptide scores from the complete and randomised databases were compared, **Figure 3.10A,** in an approach that has previously been shown to be effective for PLGS-dervived protein identifications [50]. It can be seen that the peptide scores obtained from the randomised database produce a Gaussian peak centered about a score of 4.25. The results from the complete database have an additional tail that begins at around a peptide score of 6 and represents scores from peptides almost exclusively identified from the forward database. Using these data it was possible to plot the peptide level FDR as a function of peptide score, **Figure 3.10B**. A peptide score threshold of > 5.4 was applied to the entire dataset based on the calculation of a maximum peptide false identfication rate of 4% (dotted line).

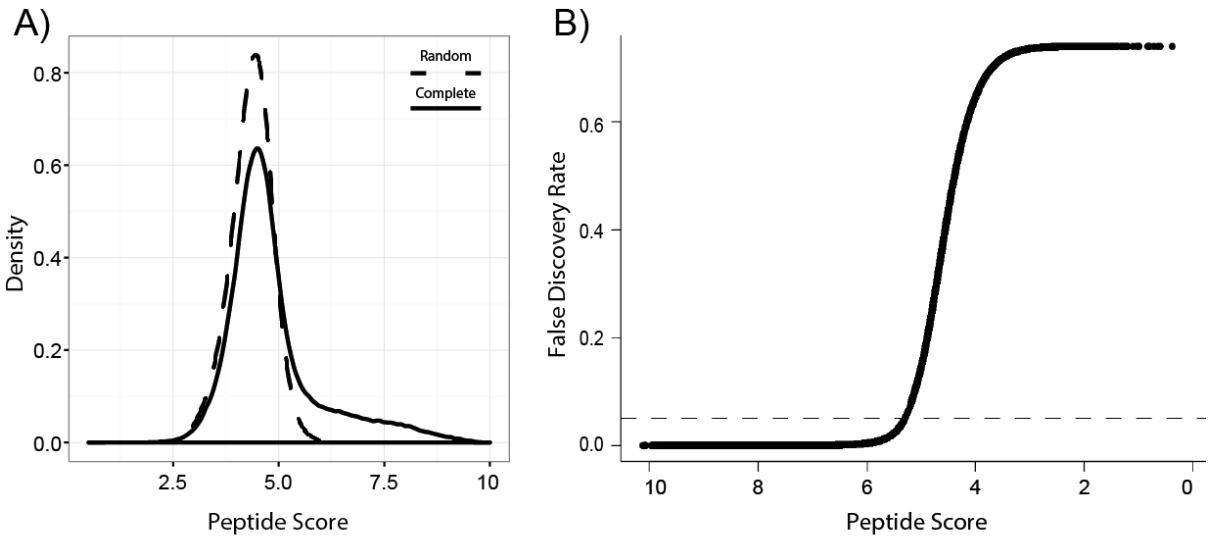


Figure 3.10 Peptide score filtering by false identification estimation using a decoy database strategy. A) Distribution of peptide score from the complete protein database (solid line) and the randomised database (dashed line). B) FDR plotted as a function of peptide score. Dotted line represents an peptide FDR of 4%

3.2.9 The Ageing Brain Proteome

Ageing is an almost universal biological process and is of interest in the context of neurodegeneration as ageing is often the biggest risk factor [51]. Proteomic studies of ageing have almost exclusively focused on the latter stages of an organism's life span [52], [53]. The initial aims of our study were to define the *Drosophila* brain proteome in relation to wild-type ageing in a more fine grained approach to previous publications. Brain tissue samples were collected at six different time points that spanned > 80 % of the *Drosophila* life span, analysed by a DIA quantitative proteomics approach and resulted in the collection of expression profiles for > 2500 proteins across the lifespan. From the protein dataset it was found <15 % were quantified with a value of zero, which is indicative of a missing value again highlighting the benefits of identification transfer in mass spectrometry-based proteomic experiments.

We took an initial unbiased approach to better understand brain proteome alterations in the context of ageing by conducting cluster analysis on proteins that demonstrated a significant fold change at any point during the time-course. A significant fold change was defined as an effect size larger than 0.7, where effect size was calculated as previously (see Chapter 2). Briefly quantitative coefficients of variation were calculated from biological repeats for each protein. CVs were plotted against the mean protein intensity and a rolling average was calculated, **Figure 3.11**. In contrast to previous data collected using a GeLCMS^{E} workflow, the data demonstrates a direct correlation between increasing protein intensity and quantitative precision. The correlation validates the use of effect size as a filter for label-free data as opposed to fold-change which fails to take into account measurement variation. Filtering successfully identified just over 1000 proteins as significantly altered across the six time points and these were taken forward for soft clustering based on the fuzzy-c algorithm [32]. The use of a soft clustering method, where elements are not confined to a single cluster, is well suited to biological time-course data where clusters are not necessarily well separated and proteins/genes can belong in multiple clusters. Clustering was implemented within the statistical environment R using the Mfuzz package. The optimal number of clusters was chosen with the use of principal components analysis, as provided by Mfuzz. Cluster numbers were varied from 5-15 and subsequent PCA plots were examined to determine how well each number of clusters represented the data. Using this method a final number of 9 clusters was chosen. Proteins with a membership score of > 0.5 were defined as being present in a cluster and z-score normalised expression profiles of proteins from each cluster were visualised, **Figure 3.12**. The functional annotation tools DAVID and Panther were used

to annotate proteins in each cluster based on gene ontology and to test for significant enrichment of any biological process, molecular function or protein classes [30], [31].

Clustering successfully identified 9 distinct proteome expression profiles and of these, 7 displayed enrichment of at least one gene ontology term therefore indicating a link between ageing and defined cellular responses. The profile of cluster 1 showed a distinct reduction in protein levels with ageing. Enrichment analysis showed the cluster to be significantly enriched for proteins that function within the mitochondrion and that are directly involved in energy production. Mitochondrial dysregulation has been previously implicated in driving the ageing process and our data indicates that this may be particularly pronounced within the brain [54], [55]. A decrease in mitochondrial function is known to lead to reduced ATP production, impaired calcium buffering and an increased generation of reactive oxygen species. Interestingly our data suggests that for a number of proteins involved in mitochondria energy production this decline does not occur only at the later stages of the *Drosophila* lifespan but is instead a gradual process that occurs throughout the lifespan and at a steady rate. In addition to mitochondria-related protein enrichment in ribosomal proteins was also noted. A reduction in ribosomal protein and transcript levels in relation to age has previously been reported in *C. elegans* suggesting that this reduction is conserved across species [56]. Within cluster 1 a large number of proteins, although displaying consistent downregulation, do not display large fold changes. Less than 7 % of proteins have a measured downregulation, when comparing T1 to T6, of more than 2-fold. The benefits of a clustering-based approach to probe global changes is therefore apparent in this instance where many of the changes detected by clustering are too small to validate with traditional statistical testing. Of the larger fold changes detected a number are linked to interesting biological functions including No extend memory (nemy) an oxidoreductase expressed predominantly in brain mushroom bodies that has direct effects on memory and locomotory behaviour [57]. Interestingly nemy is not the only highly downregulated protein that is implicated in memory and neuronal function. Numb-associated kinase (Nak) is found to be almost 4-fold less abundant at T6 and is involved in a number of neuronal specific biological processes including dendrite morphogenesis and the regulation of clathrin-mediated endocytosis [58].

Cluster 3 represents another set of proteins that display a consistent reduction in concentration in relation to the first time point, T1. The set of proteins are, however, found to drop in concentration

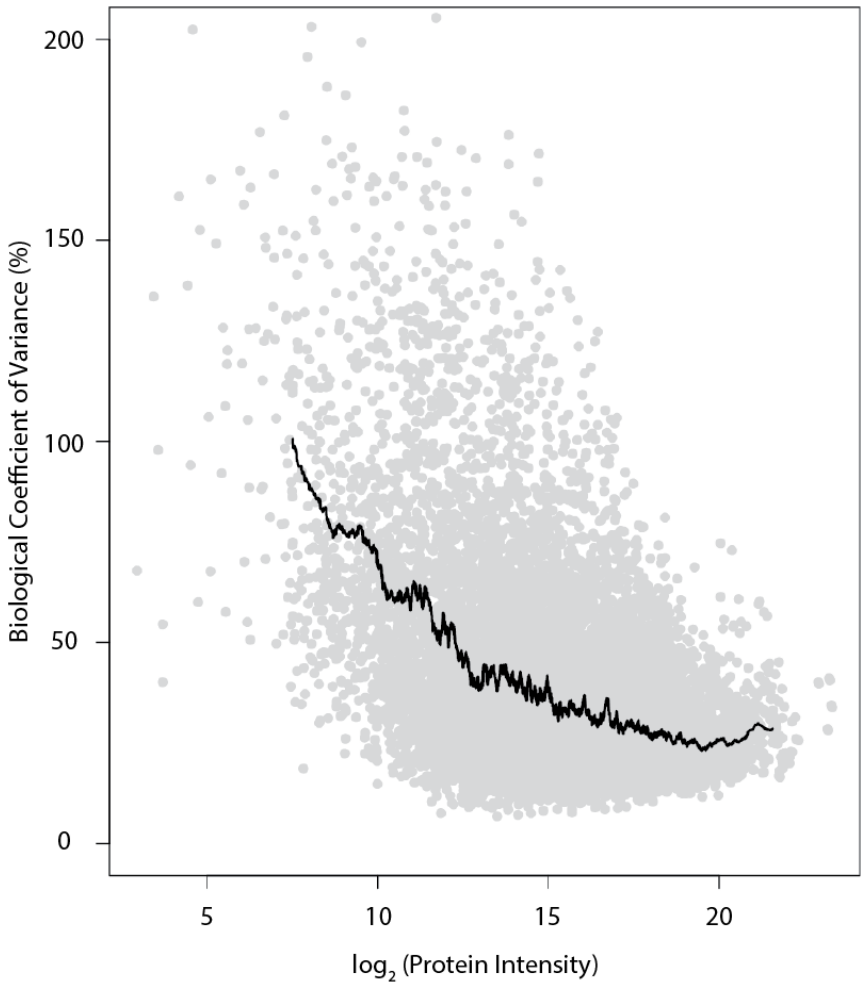


Figure 3.11 Scatterplot of quantitative biological variation as a function of protein intensity. The rolling mean of biological CV is represented by the black line

between the two earliest time points and remain at this level for the remaining lifespan. These proteins are of particular interest as the observed decline in protein levels occurs long before any ageing phenotypes are typically present and could therefore represent the earliest stages of the ageing process. Enrichment analysis found the cluster to be enriched in a number of different processes including sexual reproduction, signal transduction and the ubiquitin proteasome pathway.

Sexual reproduction in *Drosophila* is known to have direct links to longevity with an underlying trade-off between lifespan and early fecundity [59]. The trade-off is thought to be governed by negatively pleiotropic genes based upon the evidence that many of the mutations that extend lifespan also negatively affect fecundity. An early decline in sexual reproduction levels is therefore not unexpected in a *Drosophila* line selected for use in ageing studies due to maximal longevity. More interesting is

the decline in a subset of proteins involved in the ubiquitin proteasome system which is contrary to a recent *C.elegans* ageing study which found a consistent increase in the proteasome subunit levels [56]. Distinct differences between the two studies could however account for the disparity in results. The study by Dirk M. Walther and colleagues examined the global protein changes across all tissue types rather than our brain specific profiling approach. Indeed it has been previously found that

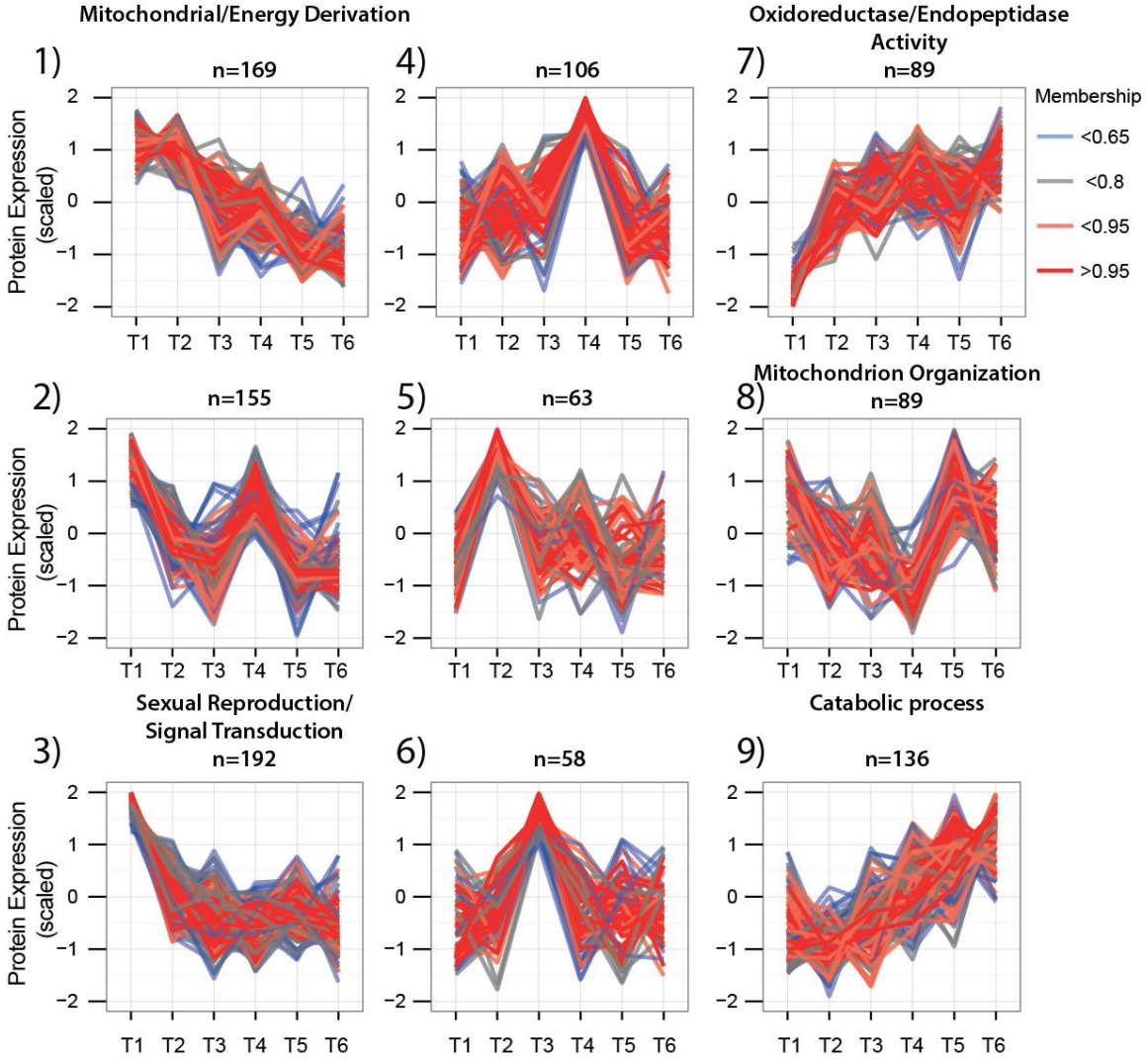


Figure 3.12 Clustering of time-course expression levels from proteins that were found to display a significant expression change (effect size >0.7) across the data. Levels were clustered using the fuzzy c-means algorithm and colours indicate the cluster membership score for each protein. Labelled clusters indicate significant enrichment (EASE score (modified Fischer Exact p Value) < 0.005 or FDR < 0.05) of a biological process or molecular function. Enrichment analysis was performed in DAVID and Panther

the activity and expression levels of the proteasome complex within the central nervous system of rat decline as a result of ageing are thought, in part, to be due to the high levels of reactive oxygen species produced within the brain [60]. Unlike cluster 1, cluster 3 contains a far larger number of proteins that decrease by 2-fold or more (36 %) with up to a 33-fold decrease in the levels of Aldose-1-epimerase, a protein directly involved in glycolysis and gluconeogenesis. Within this cluster subset reside proteins directly involved in neurogenesis and central nervous system development and could suggest the presence of neurogenesis in the early stages of the adult brain which is reduced or arrested at an early age. Until recently it was not thought that the formation of new neurons occurred within the adult *Drosophila* brain but a recent paper by Fernandez-Hernandez and co-workers suggest a level of plasticity that might provide a mechanism for brain regeneration [60]. Our results provide an insight into why levels of neurogenesis might be low in the adult fly brain.

Conversely we were interested in proteins that were consistently upregulated as a result of ageing, represented by clusters 7 and 9. Both clusters contain fewer proteins than their downregulation counterparts but were still found to be enriched for gene ontologies. Cluster 7 contains proteins that are almost exclusively upregulated between T1 and T2. The clusters were enriched in proteins that act as either an oxidoreducatase or dehydrogenase. Of these proteins, CG30375 is highly upregulated (> 20-fold) and is known to have serine-type endopeptidase activity. Cluster 9 which displays a consistent and progressive increase in protein concentration is enriched for organic acid catabolic processes. An increase in catabolism during ageing has been previously reported and is thought that this, alongside a decrease in anabolic signals, to contribute to age-related cellular degeneration through a loss of catabolic-anabolic homeostasis.[61] Further to this branched chain amino acid catabolism has previously been shown to effect lifespan in *Drosophila.*[62] Interestingly catabolic malfunction in ageing has been implicated in mitochondrial damage and would fit with our previous findings [63].

Following cluster analysis a more traditional approach to identify statistically significant protein-level changes across the six time points was taken. More established statistical tests such as the Student's t-test or linear models for microarray data (Limma), although well suited to the pairwise comparison of different biological conditions, are not designed to be used for modelling time-course data and therefore lack statistical power. Recently a new statistical algorithm, Extraction of Difference Gene

Expression (EDGE), which has been implemented as a Bioconductor package, was developed by the Jon Storey lab with strategies specifically designed for analysing results derived from time-course experiments [29]. Statistical testing was performed using the optimal discovery procedure and was implemented within the EDGE Bioconductor package. Using this powerful approach to multiple hypothesis testing, our ageing dataset was tested for differential expression. At a q value of 0.05 just over 340 proteins were determined to be differentially expressed representing around 10% of the total number of proteins successfully quantified across all time points. To validate the EDGE approach, we undertook a comparison to the more traditional time-course analysis approach of conducting an ANOVA where time is included as a variable, **Figure 3.13.**

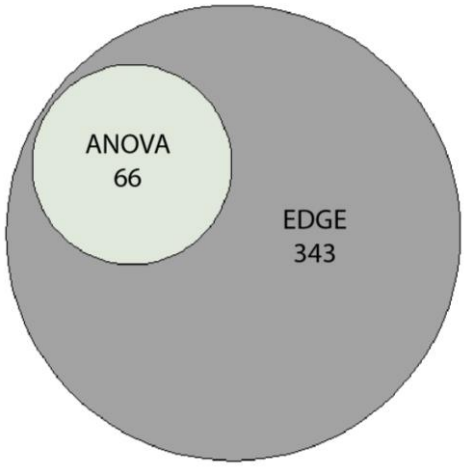


Figure 3.13 Venn diagram of the overlap between proteins identified as differentially expressed from the ageing data by an ANOVA test or by the Bioconductor package EDGE

Pleasingly EDGE was found to outperform the ANOVA test by 80% and successfully identifies all proteins identified by the ANOVA method.

Enrichment analysis confirmed the presence of a large number of proteins involved in metabolic processes, as indicated by previous cluster analysis. The large number of metabolic proteins (>140) is unsurprising considering the energy-intensive nature of neural processing. It has been previously estimated that the brain accounts for around 20% of a human bodies resting metabolism. A dysregulation of brain metabolism has long been postulated for being a cause of age-related decline and many neurodegenerative disease symptoms, AD included, have been linked to a metabolic imbalance, toxicity and dysfunction [64]–[66]. Our data confirms changes in metabolic protein levels

represent a large proportion of proteome alteration induced upon ageing. Proteins implicated in translation were also found to be enriched within the differential expressed subset. Translation rates from mRNA to protein are known to mediate life span extension in both *C.elegans* and *Drosophila*. Indeed the inhibition of mTOR, a major signalling pathway in ageing research, increases longevity in part by reduced protein synthesis rates [67], [68]. Interestingly a proportion of the regulated proteins, including four translation factors and ribosomal subunits, decrease in concentration in relation to age. Previous cluster analysis had indicated this for ribosomal proteins but did not find enrichment in translation factors. Protein synthesis levels have been previously found to decline in respect to increasing age and the downregulation of translation factors Ef1-beta, EF1-delta and eIF-5A all of which are involved in translation elongation is indicative of this phenotype.

To allow a more global view of differential expression in respect to increasing age, statistically significantly altered proteins were queried for protein-protein interactions within the STRING database [33]. All proteins confirmed to interact with at least one other protein, of medium confidence or higher, were retained for visualisation within Cytoscape. Mean abundance ratio values, in reference to T1, were calculated for each of the 5 time points and were represented by node colour within the network. This allowed the visualisation of the PPI network expression levels with respect to time, **Figure 3.14 and Appendix E**. A total of 263 proteins were found to form an extensive network with the visible presence of multiple sites of highly coordinated interactions. Interrogation of the largest of those sites confirmed results from previous analysis indicating a systematic decrease in translation-associated proteins. PPI network analysis was also able to expand upon previous translation control alterations that were occurring upon increasing age. Within the network a large number of mRNA binding proteins including mapmodulin, hoip, rump and mub were found to interact or cluster close to ribosomal proteins. Perhaps more significantly was the presence of proteins that are directly implicated in chromatin remodelling and the Wnt signalling pathway. A loss of function of the Wnt signalling pathway has been previously proposed to lead to the onset of many of the hallmarks of AD and has been linked to a number of other diseases including cancer.[69] The pathway is a key activator of neuronal development but the regulation of the pathway in relation to age is not well understood. Our data suggests a differential regulation of the pathway which could have implication for gene transcription and cell fate.

The second site of high PPI coordination was found to correlate with the high levels of mitochondrial and metabolic enrichment we observed during cluster analysis. Again we observed a downregulation across each time point for the majority of proteins found within this site. Functional analysis of the proteins within this cluster identified NADH dehydrogenase activity to be a common biological process. The differential expression of this set of proteins, altered levels of COX6B and altered cytochrome-c oxidase activity is strong evidence for the potential dysregulation of the electron transport chain with respect to age.

As previously mentioned, recent work by DM Walther and colleagues from the Mann and Hartl labs was published on a comprehensive proteomic time-series analysis of ageing in *C.elegans* [56]. Although there are significant differences between the aims and experimental workflow of the Mann and Hartl labs and our study we thought it of interest to examine whether any of the findings of the study were validated in our data, **Figure 3.15**. The work by DM Walther and co-workers finds evidence for a dysregulation of the proteostasis network with a downregulation of ribosomal subunits and an upregulation of proteasome constituents. Our data confirms that an average reduction in ribosomal levels is conserved in *Drosophila* brains, however, the measured reduction is less pronounced and for a small number of subunits we actually observe a large increase in abundance, **Figure 3.15A**. Proteasome levels on the other hand show the opposite trend to that seen in *C.elegans* tissue with a downregulation in our data, **Figure 3.15B**. More interestingly we measure a large amount in variation in proteasome subunit abundances, a trend also seen the *C.elegans* study, and which might provide evidence for a dysregulation of the proteasome rather than a more general reduction in proteasomal activity. In **Figure 3.15C** it can be seen that our data agrees well for detected HSP70 and HSP90 chaperone proteins, which remain largely unchanged across the lifespan. An exception is found for Hsc70-interacting protein (HIP) which is significantly and consistently downregulated with increasing age ($q < 0.0005$). Evidence suggests that HIP through its own chaperone activity, where it stabilises the ADP state of Hsc70, contributes to the chaperone activity of Hsc70 [70]. Finally proteins involved in oxidative stress protection, were found to show no general trends upon ageing whereas the Mann and Hartl labs measured increases for two oxidative stress defence proteins. Where the two data might successfully validate a large age-related change is

for superoxide dismutase [Cu-Zn] (SOD-3). The current study measures a large reduction in protein

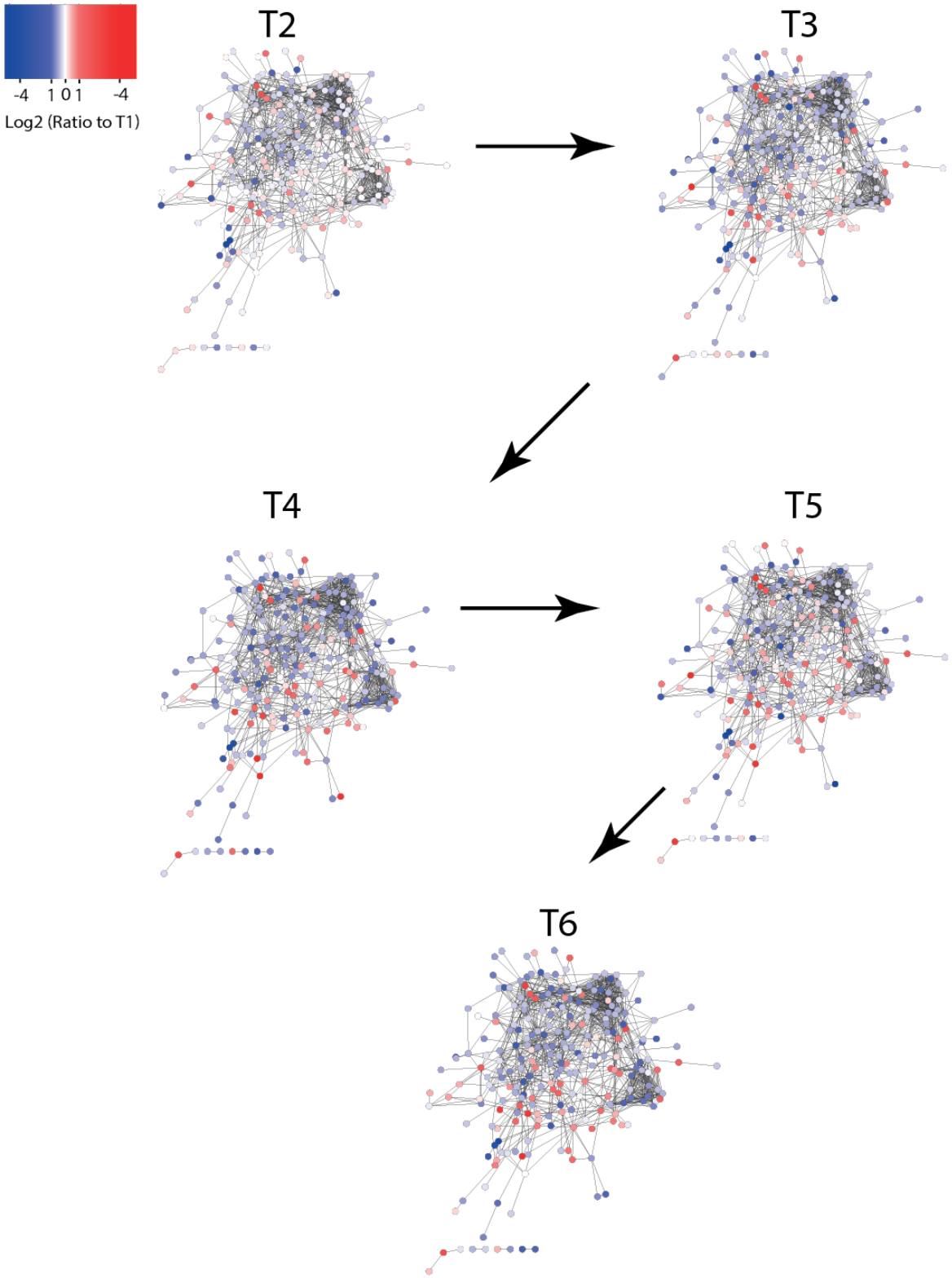


Figure 3.14 Longitudinal protein-protein interaction network of brain specific differentially expression across the *Drosophila* lifespan. PPI's included are as defined in STRING as medium confidence or higher. Expression data is represented by node colour and was visualised in Cytoscape

concentration at T4. This reduction remains at least 2-fold less in comparison to levels at T1 for the duration of the lifespan. The data from the Mann and Hartl labs is less conclusive with no change observed up to 12 days after which the protein is no longer successfully quantified. The loss of detection after day 12 is likely to be due to protein levels being reduced below the instrument limit of detection. The data therefore can be taken to be indicative of a very large reduction in SOD-3 levels after day 12. SOD-3 has been previously implicated in longevity, through enhanced resistance to oxidative stress and a decline in levels with increasing age could contribute to increased mortality rates [71], [72]. Across the entire two sets of data we observed little overlap between proteins detected with an observed fold change > 2, although this does represent a statistically significant overlap ($p < 0.0001$), **Figure 3.15E**. This is perhaps not surprising considering that the data obtained by the Mann and Hartl labs was collected from whole worms and quantitative data, is therefore, an average value for all tissues present in the organism. Higher abundance tissues are likely to be better represented with this approach. As worms have a limited central nervous system (302 total neurons), neuronal specific expression changes are likely to not be fully represented in the Mann and Hartl labs dataset.

3.2.10 Brain Proteome Alterations Resulting from Chronic Abeta Induction

The commonly accepted aetiology of Alzhiemer's disease is an increase in production and/or the accumulation of Aβ42 peptide which causes peptide oligomerization and the formation of extracellular aggregates. It was the initial discovery of extensive brain plaques in AD individuals, which were composed primarily of Aβ42, which led to the amyloid cascade hypothesis being proposed.[73] The hypothesis describes the downstream events following Aβ accumulation and deposition, resulting in synaptic and neuronal dysfunction. In the 10 years since the discovery, research into AD has focused on the effects of Aβ peptide aggregation. The approach taken has been primarily to study animal models, which, including *D. melanogaster*, either directly express Aβ42 or express genetic variants of PS1 and PS2 that indirectly produced Aβ peptide species through altered APP processing. In our study we measured brain proteome changes in an adult-onset AD *Drosophila* model that encompassed the majority of the lifespan. Initially, we were interested in defining proteins that displayed expression changes in the presence of Aβ42 induction and to group them based on their

profiles. As with the ageing data a fuzzy-c clustering algorithm was used, after filtering for proteins that had an expression change across the time-series with an effect size larger than 0.7, **Figure 3.16**. As previously the optimal cluster number was determined by varying the allowed cluster numbers between

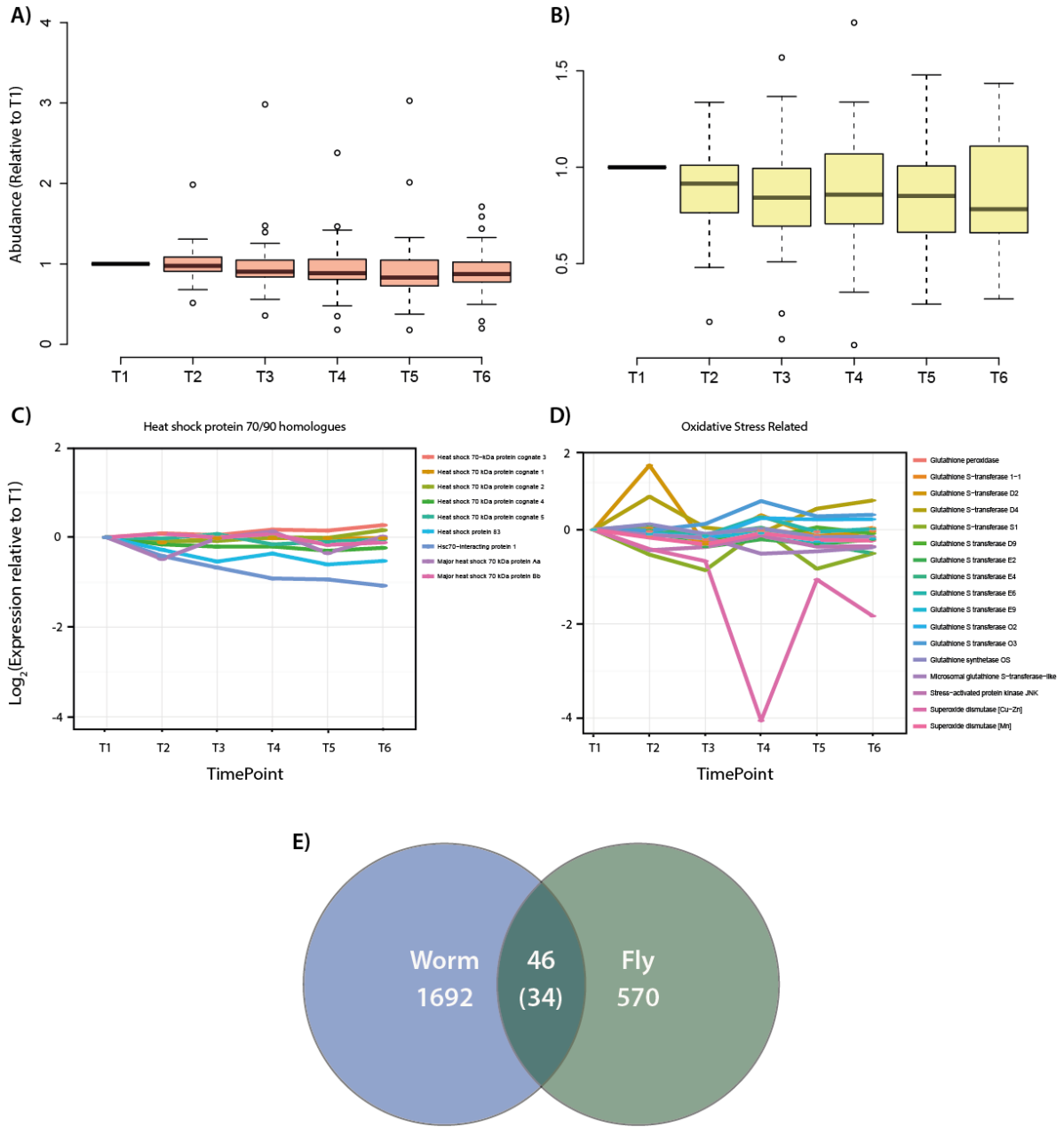


Figure 3.15 Comparative analysis of ageing data with the results published by DM Walther and co-workers studying the ageing in *C. elegans.* **A)** Abundance changes of ribosomal proteins. **B)** Abundance changes of proteasome subunits. **C & D)** Abundance profiles of HSP 70/90 homologues and oxidative stress related proteins. **E)** Venn diagram of the overlap between the two data set including only proteins with a measured fold change of >2. Number in brackets indicates the number of *Drosophila* specific proteins

5-15. Resultant PCA plots were then visually inspected. Using this method an optimal number of 7 clusters was determined, two less than used for the ageing study. The lower number of clusters for these data is unsurprising considering the lower degree of freedom afforded from a four point time-course. Initially apparent from the Aβ data clustering was that the data showed less protein to protein abundance variation than the ageing data, with each cluster having a more coherent expression profile. It was possible to quantify the visual trend by comparing the average membership score within each Aβ cluster to the ageing data clusters. Membership scores from the Aβ-induced clusters were found to be significantly higher ($p < 0.00001$) than those from the ageing data with an average score of 0.93 and 0.87 respectively. The cellular response to an extensive protein aggregation event, as is seen upon Aβ induction, could be expected to be more consistent than those changes characterised by the ageing process, a process known to be complex and driven by a multitude of deleterious mechanisms. Our data, therefore, is suggestive that our experimental protocol has identified protein abundance changes that are the direct result of Aβ42 induction toxicity. Although a large amount of overlap is observed between the cluster enrichment analysis for the ageing and Aβ data, differences present may provide information on the cellular response to Aβ aggregation. Cluster 2 follows the same profile as Cluster 1 from the ageing expression, with a consistent reduction in protein abundance across the lifespan. Similar to that seen in the ageing data the cluster is enriched in mitochondrial proteins, specifically those involved in the electron transport chain. Such an overlap is unsurprising considering a reduction in mitochondrial function has been reported in both ageing and Alzheimer's disease previously [55], [74]. Interestingly cluster 2 is also found to be enriched in actin proteins that form cytoskeleton structures, an enrichment term not observed in any of the age-related clusters. Cytoskeletal abnormalities have been reported to be present in AD brains and are thought to be associated with dystrophic neurites, where amyloid plaques induce a transformation of neuronal cytoskeletal proteins [75]. Indeed tau is a major neurotoxic aggregating protein that is implicated in the later stages of the amyloid cascade hypothesis [76]. Reduction in cytoskeletal abundance may be representative of an increased formation of dystrophic neurites that are likely not to be solubilised by our protein extraction method. A major difference between the two clusters from the two respective time-course datasets were the number of proteins detected with a large reduction (> 2-fold) in abundance in the presence of Aβ peptide. Notably the protein CG5334, a polyubiquitination

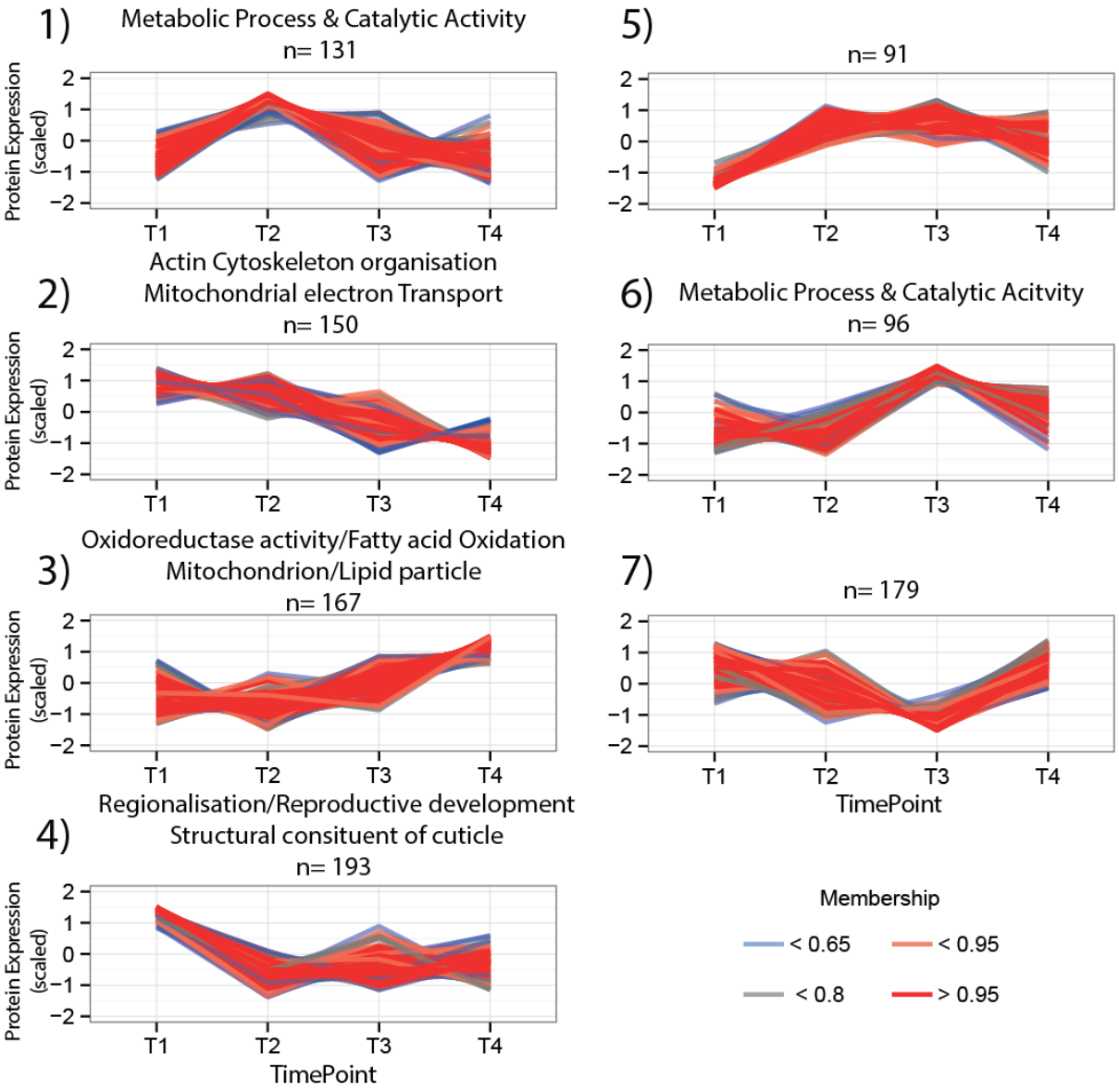


Figure 3.16 Clustering of expression profiles for Aβ induced time-course. Only included are proteins that had an expression change with an effect size > 0.7. Labelled clusters indicate a significant enrichment (EASE < 0.005 or FDR < 0.05) of a gene ontology as determined by enrichment analysis with DAVID or Panther respectively

associated protein, had reduced in abundance by over 100-fold by T4, whilst pck which aids in the establishment of the glial blood brain barrier had a 9-fold reduction across the time-series.

Cluster 3 conversely contains upregulated proteins many of which are involved in fatty acid oxidation, a process which is closely tied to AD pathology [77], [78]. Interestingly increases in abundance for the majority of proteins are not observed until T3, a point at which the mortality rate is soon to start to rapidly increase. This suggests that the cluster may highlight proteins whose abundance change either contributes or is a compensatory mechanism to an increase in mortality. Maximum protein

levels are quantified at T4, the point at which mortality is at its peak, **Figure 3.3A**. It is thought that increasing levels of oxygen radical formation in mitochondria is present in AD brains due to fatty acid breakdown, which are susceptible to free radical damage in a process called lipid peroxidation [79]. Our data demonstrate a correlation between fatty acid oxidation levels and mortality rate which could help validate this theory.

Both metabolic processes and catalytic activity were found to be enriched in two clusters, both of which showed variable abundance levels across the time-course. This could be due to the dysregulation of different metabolic pathways and indeed previous metabolomics studies have identified the perturbation of numerous different pathways in the cerebrospinal fluid taken from AD patients [79].

Differential expression was, as previously, calculated using the EDGE package. The same settings as those used for statistical testing of the ageing data were used excluding the degree of freedom which was set to 1 less than the number of time points measured, 3. Using this method, 124 proteins were identified as having a statistically significant abundance change across the time-course ($q < 0.05$). We initially wished to confirm that our data captured the cellular response to Aβ peptide toxicity by using the ageing data as a reference. We therefore examined the overlap between differentially expressed proteins identified in each of the two data sets, **Figure 3.17**. Little overlap was observed, with 124 out of a total of 141 proteins found to be specific to the Aβ data set. This indicates that the majority of proteins identified as differentially expressed from the Aβ data are either a direct result of Aβ toxicity or expression changes not detectable in our ageing data are exacerbated by Aβ expression to a detectable level.

Gene ontology analysis of differentially expressed proteins found that a large number were involved in metabolic processes, confirming the results from cluster analysis (59). One proposed mechanism for neurodegeneration in AD is the onset of brain metabolic dysfunction and indeed metabolic profiling of AD brains, primarily with the use of LC-MS has identified the dysregulation of various metabolic pathways [80], [81]. Within our data there were no highly represented pathways suggesting a non-specific effect on metabolism. There was however a number of dehydrogenase (14) and transferase (11) related proteins. Of the dehydrogenase proteins, aldehyde dehydrogenase was found to be upregulated with age ($q = 0.0005$) with its highest abundance measured at T4. Interestingly the same

upregulation was observed in the ageing data but with a larger fold change which initially suggested that the observed expression change is most likely to be as a result of ageing. Mitochondrial aldehyde dehydrogenase has however has been previously linked to AD, where it has been observed that a deficiency is a significant risk factor for late-onset AD [82]. It is also known

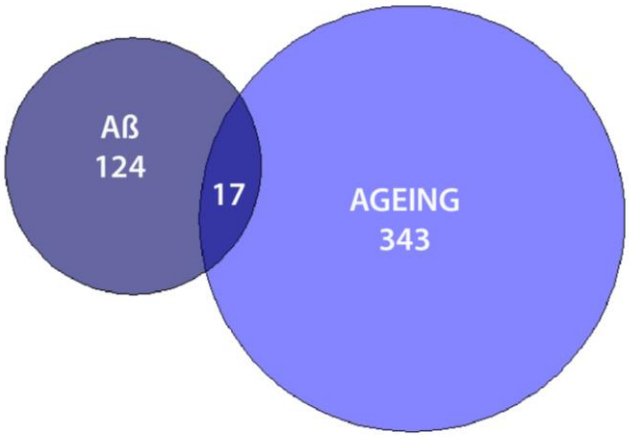


Figure 3.17 Venn diagram of the overlap of differentially expressed proteins identified from the ageing data and the Aβ induction data

that aldehyde dehydrogenase is protective against oxidative stress, a proposed mechanism of pathogenesis in AD. When directly comparing protein levels between the two datasets no statistically significant difference was measured, indicating that differential regulation is not linked to Aβ toxicity in our model. The data, therefore, may be indicative of an upregulation as a protective mechanism against increasing oxidative stress as a result of ageing instead of Aβ aggregation but one which may incidentally protect against Aβ induced oxidative stress in AD. Other metabolic proteins to note due to large abundance alterations were the protein scaf, a serine protease (30-fold downregulated), rhi a transcription factor (30 fold downregulated) and Eukaryotic translation initiation factor 3 subunit L (8 fold downregulation).

Also highly enriched (FDR < 0.001) were proteins with oxidoreductase activity, where a large overlap with proteins also having dehydrogenase activity was observed. Within this subset the enzyme cytochrome P450 4e2 was quantified at 5–fold less after T1 ($q = 0.04$), an effect not observed in the ageing data. Interestingly it is known that cytochrome P450 enzymes metabolise a wide range of

exogenous and endogenous compounds and are found in a range of tissue types. It is thought that alterations in cytochrome P450 metabolism could regulate brain behaviour, disease pathology and treatment outcome through altered drug metabolism [83]. Decreased levels of P450 4e2 could likely contribute to neurotoxicity through decreased clearance of toxic compounds which would otherwise have been cleared. A similar hypothesis has been proposed for the link between pesticides and the risk of PD, where farmers with a Cytochrome P450 2D6 variant that metabolises compounds at a slower rate were found to be at an increased risk of PD.[84] Global alterations in oxidoreductase activity could contribute to changes in detoxification pathways thereby increasing neuronal susceptibility to damage.

In a similar approach to the previous data, differentially expressed proteins were queried in STRING to identify protein-protein interactions and protein subsets that might be performing similar cellular functions. Although the resultant network was not as expansive as that observed from the ageing data, with the use of ClusterViz, we were able to identify at least four distinct clusters of PPIs, **Figure 3.18**. Cluster I, labelled in red, is the largest of the clusters and contains proteins with a range of cellular functions including neuronal stem cell maintenance, locomotory behaviour and actin cytoskeletal binding, corresponding to the proteins Trehalase, Tdc2 and Moesin/ezrin/radixin homolog 1 respectively. Well represented in this cluster are oxidoreducatase related proteins including Sodh-2, CG2064 and CG1548. Cluster II, labelled in yellow, is highly represented by proteins either involved in the maintenance of protein folding or the control of protein translation. HSP23 and 27 are both identified as downregulated across the time-series, but interestingly HSP23 is also downregulated with age. This taken with the identification of age-related expression changes of HSP27 in previous experiments (Chapter1) indicates that these changes are for the most, induced by ageing. A reduction in chaperone concentrations to aid protein refolding could, however, have dramatic implications for a misfolding disease such as AD. Another protein closely linked with protein folding is ref(2)p, which is found to be 3-fold upregulated, and is involved in both ubiquitin-dependent protein catabolism and the positive regulation of macroautophagy. Interestingly ref(2)p has been shown to be a key component in the formation of protein aggregates in *D. melanogaster* models of neurodegenerative diseases.[85] Our results therefore fit well with this model. Two other proteins within the cluster were found to be common between the two datasets, elav and banacal. Elav is a neuron specific mRNA binding protein required for central nervous system maintenance and is used in the Elav-GS expression system

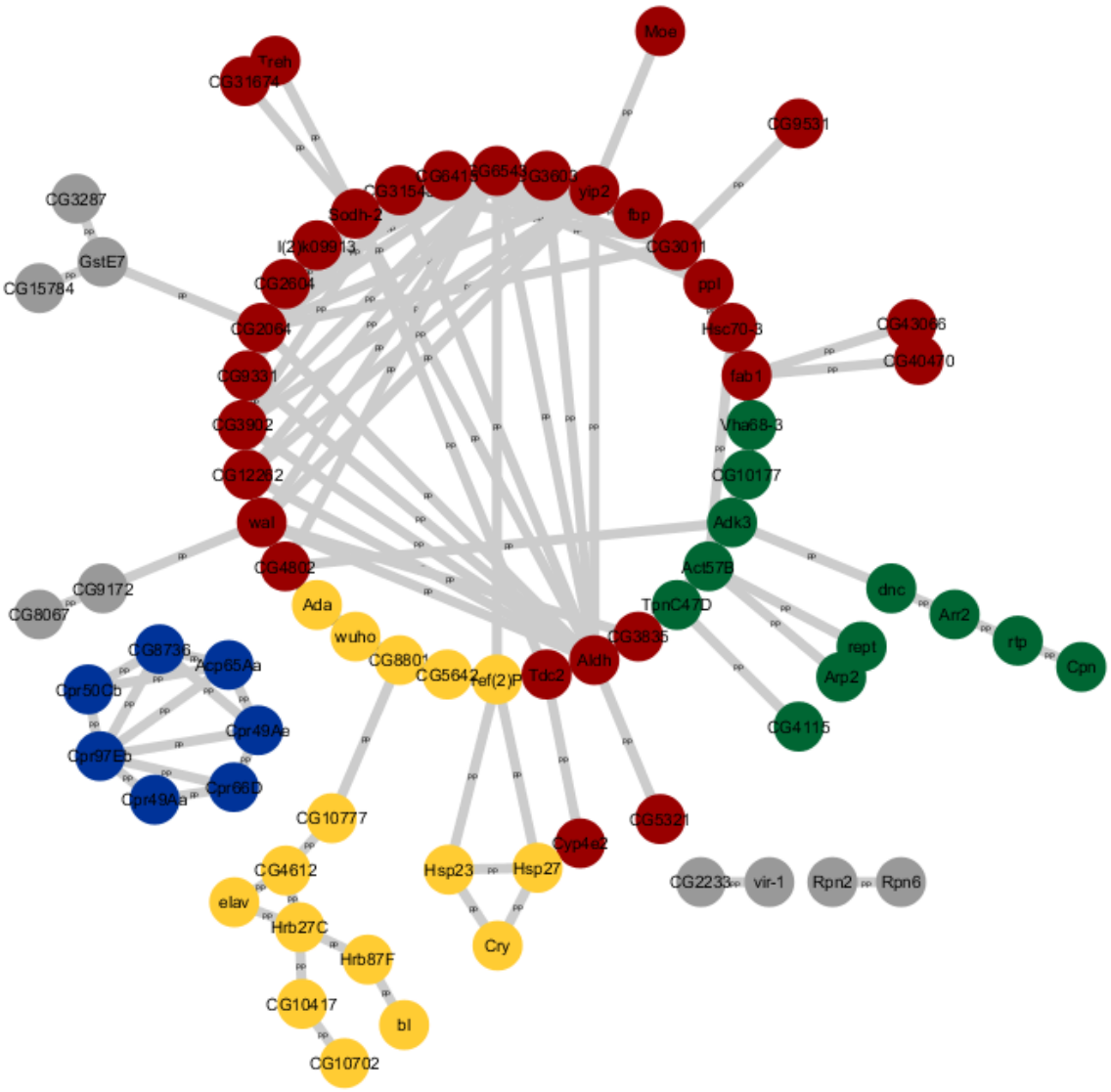


Figure 3.18 Protein-protein interaction network of proteins identified as differentially expressed across the Aβ induced time-course. Interactions are those defined as of medium confidence or higher as defined with STRING. Clusters (as labelled by colour) have been calculated using the EAGLE algorithm from the ClusterViz application and as applied within Cytoscape

[16], [86] Reduced amounts are seen across both time-course datasets and are likely a signal of neuronal death in both cases. The expression profile for bancal is similar to elav, with a reduction seen across both time-series. Bancal is a transcription factor binding protein that is linked to cell fate and proliferation.[87] Elevated levels of bancal and other heterogeneous nuclear ribonucleoproteins (hnRNPs) have been previously correlated with improved neurodegenerative phenotypes in flies.[88] Reduction as a result of ageing could therefore contribute to increased vulnerability to Aβ toxicity. All other proteins within the interaction network were found to be specific to the Aβ data and are almost

exclusively involved in the regulation of transcription/translation, mainly by mRNA or tRNA interaction. A reduction in translation rates has been linked previously to neurodegeneration in prion disease via eIF2-α activation, however, little has been published in the context of AD.[89] Interestingly, however, neuronal death in AD has been proposed to as a result of ER stress and dysfunction.[90] The unfolded protein response (UPR) is major ER stress mechanism and causes an overall reduction in mRNA translation.[91] Our data in addition to the identification of a downregulation of eukaryotic translation initiation factor 3 subunit L (eIF3l) demonstrate a degree of protein synthesis control in response to Aβ induction which could be attributed, in part, to increased ER stress. Cluster III, labelled in pink, contains proteins involved in signal transduction, learning and rhabdomere development and therefore represent alterations in brain and eye function as a result of protein aggregation. Two actin related proteins are also present, Actin-57B and Actin-related protein 2, both of which regulate growth at the neuromuscular junction. Interestingly both are downregulated 2-fold across the series, suggesting dysregulation at the neuromuscular junction. Finally Cluster IV contains cuticle and chitin cuticle-related proteins that are the major structural constituents of the exoskeleton that surrounds the *Drosophila* brain. Although these could be viewed as contaminant proteins that were not effectively removed during dissection, the consistent nature of their downregulation suggests that the results may have some biological relevance. Alongside this, the identification of only one differentially abundant cuticular protein from the ageing dataset demonstrates an Aβ specific response. The differential abundance of a set of cuticular proteins has not to our knowledge been previously reported for fly models of neurodegeneration but could be hypothesised to be caused by the degeneration of brain tissue and therefore the association between brain tissue and cuticle.

Another approach taken to prioritise differentially expressed proteins was to compare our data to data obtained from a similar but orthogonal approach conducted at the transcript level. The data as collected by G. Favrin and colleagues compares both age-matched and survival matched flies either absent of Aβ (control) or chronically expressing the peptide [92]. Here three time points were analysed per experimental design. It is a reasonable conclusion that an expression change measured at both the protein and mRNA level from experiments performed at different times and by different labs can be confidently assigned as a cellular response to Aβ misfolding. A total of 29 differentially abundant proteins from the protein dataset were found to overlap with genes identified as significantly

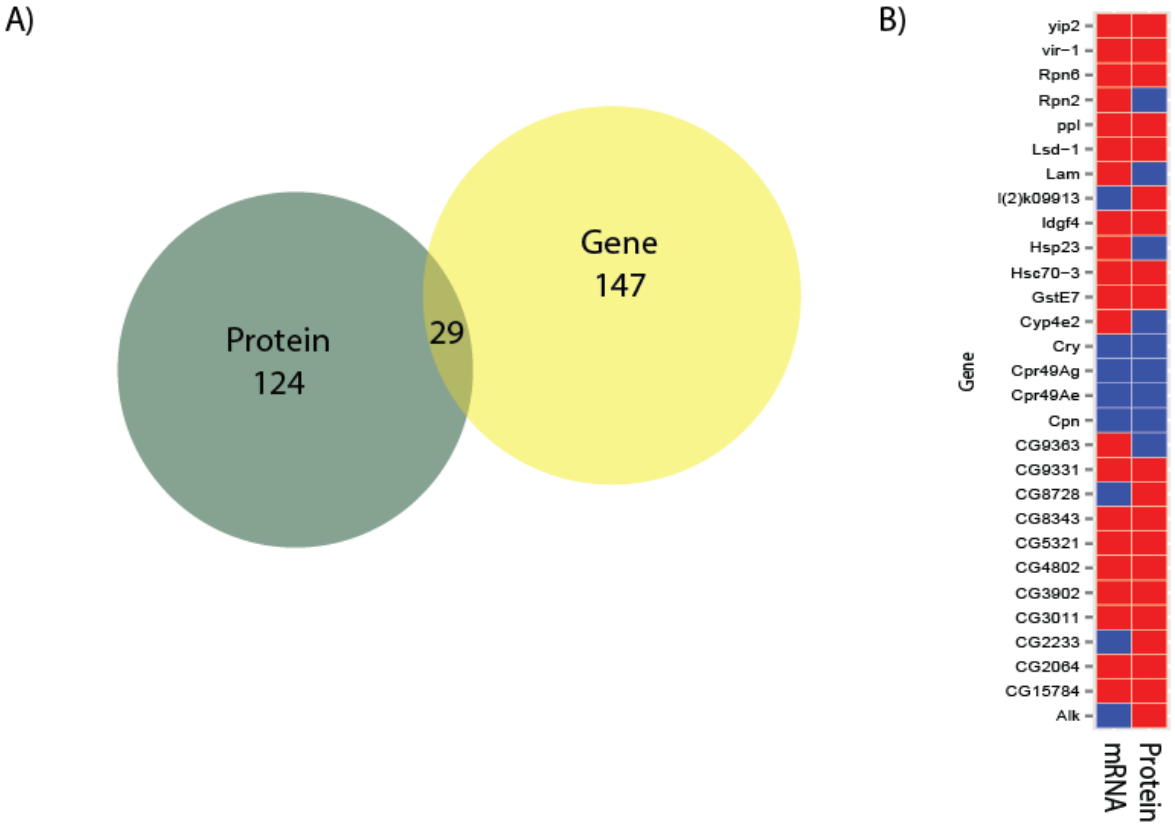


Figure 3.19 Comparison of differentially expressed proteins identified in the Aβ time-course proteomic analysis to genes found to be regulated in microarray analysis by G. Favrin and co-workers in their Aβ time-course data. **A)** Venn diagram of gene overlap. **B)** Heatmap of direction of regulation (for mRNA and protein) for genes common between the two dataset. Red indicates an upregulation and blue indicates downregulation

altered in the microarray dataset, **Figure 3.19A**. Taking into account the size of the average *D. melanogaster* microarray gene set (18,500) and the number of differentially expressed proteins identified in the two datasets this overlap represent a statistically robust overlap ($p < 4x10^{-34}$) and successfully validates our approach for prioritising proteins of interest based on prior unbiased experiments. We were subsequently interested in the proportion of common identification where regulation at the gene and protein level was analogous, **Figure 3.15B**. Pleasingly it can be seen that for the majority of genes both mRNA and protein abundance are regulated in the same direction in the presence of Aβ peptide expression. This would indicate for those genes, protein abundance is controlled mainly by transcriptional regulation. These include the upregulated lipid storage droplets surface-binding protein 1 which is involved in lipid storage and neurogenesis and the downregulated calphotin protein that is key to rhabdomere development and photoreceptor cell survival [93]. Proteins where mRNA and protein levels do not correlate could potentially represent a situation where protein

abundance levels are regulated at a post-transcriptional level through, for example, protein degradation or other known regulators of translation rates. Validation of these results with orthogonal techniques such as immunoassay assay quantification and quantitative polymer chain reaction (qPCR) is required to confirm this hypothesis.

Finally, although some of the proteins did feature in previous analyses, proteins common between the ageing and Aβ data were examined for their function and regulation. We were interested in identifying whether these shared proteins exhibited the same behaviour across the two time-series and whether the degree of abundance changes were similar, **Figure 3.20A**. Similar expression changes across

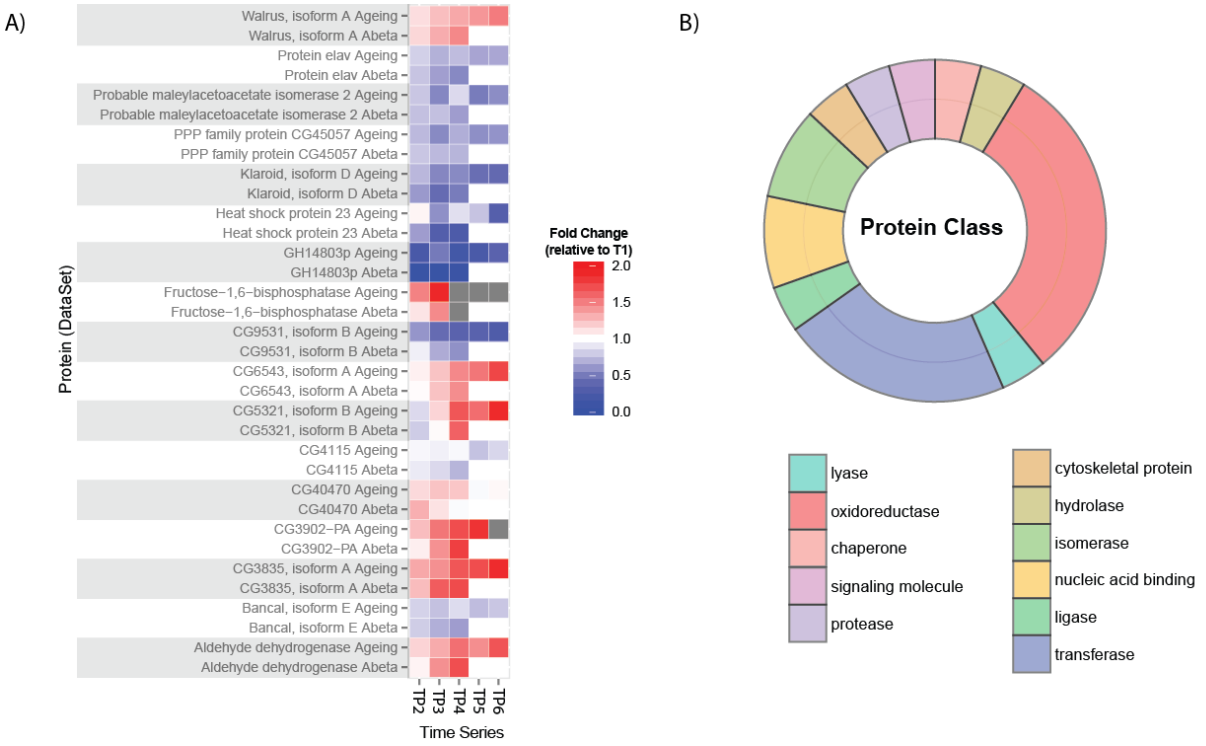


Figure 3.20 Comparison of proteins identified as differentially expressed in both the ageing and Aβ time-course. **A)** Heatmap of fold change at each time point (in reference to T1).**B)** Protein class (Panther) analysis

time were observed in each of the datasets. Notably, when comparing samples obtained at the same time point, abundance alterations measured in the Aβ induced samples were almost entirely more pronounced than in the ageing samples. Indeed T4 abundance from the Aβ time-course correlates better, in terms of fold change, with T6 abundance from the ageing time-course than it does with its respective counterpart. These data indicate that, at least for this protein subset, the aggregation of Aβ speeds up cellular processes observed during ageing. The common proteins were found to belong to

a number of different protein classes as defined by Panther, **Figure 3.20B**. As previously observed during cluster analysis, the most well represented classes were oxidoreductase and transferase activity. This may reflect shared response to increases in oxidative stress which is known to contribute to neurodegeneration in both AD and normal ageing [94].

3.2.11 Characterisation of Aβ Toxicity in Reference to Age Matched Controls

A third approach taken to extract information from our quantitative time-series was to directly compare protein abundances measured from the two time-courses. This allowed us to identify differentially expressed proteins upon Aβ induction with reference to an exact age-matched control. Prior to the comparison, data were processed and normalised together in order to reduce the contribution of any global differences between the samples. This meant, therefore, that we focused on only the most confident protein abundance differences in our analysis.

Data at each time point were compared using a two sided Student's t-test with unequal variance. Statistical tests were performed on $\log_2$ transformed Hi3 protein intensities extracted from Progenesis post alignment, normalisation and protein grouping. To reduce the false discovery rate differential expression was defined as a protein having a P value < 0.05 in a minimum of two out of the four time points tested. Using these criteria minimal differences were observed between the two datasets with 44 proteins identified as having a statistically significant abundance difference between the two conditions. The cause for these smaller measured differences may be two-fold. Firstly the use of a t-test is likely to reduce the power of the analysis in comparison to the previous approach used [29]. Secondly it may be that the majority of expression differences between the two conditions are smaller than is possible to confidently distinguish with the chosen number of biological repeats. If this is the case, as is often found in transcriptomics, enrichment analysis (based on alterations in abundance) can indicate whether small expression changes are common to biological processes. To test for this, enrichment analysis was performed (Panther) using calculated ratios of Aβ vs age matched control at T4. This was based on the assumption that upon chronic induction of misfolded peptide the most pronounced differences between the two conditions are likely to occur at the latest time point. After Bonferroni correction no gene ontologies were found to be significantly enriched. Our data therefore indicates that for the set of proteins quantified by our methodology there are no processes that are

commonly regulated in response to Aβ misfolding, when compared to normal ageing controls. Uncorrected p values obtained for enrichment analysis were subsequently considered as an approach to prioritise differentially expressed proteins. The biological process, response to endogenous stimulus, had the lowest p value ($p = 0.0045$, upregulation) and was based on ratio values quantified from 4 small GTPase proteins. This was of particular interest because of its relation to previous findings (Chapter 2) where a set of GTPase proteins were found to be differentially expressed in aged flies expressing Aβ peptide. Although the proteins identified from enrichment analysis are not those identified in previous work it may be further evidence of a common dysregulation upon Aβ misfolding. Other upregulated pathways ($p < 0.05$) were three different metabolic pathways, phospholipid, monosaccharide and cyclic nucleotide, which collectively accounted for 75 proteins.

Differentially expressed proteins were manually queried in Uniprot and Flybase for annoted biological processes which resulted in the identification of a number of proteins that performed cellular functions that were known to be directly relevant to AD based on previous publications, **Figure 3.21.** Expression of Hsc70-3 is observed to increase significantly as a result of Aβ misfolding in three out of the four time points, **Figure 3.21A**. Interestingly this trend is similar to that observed in the previous proteomic study (Chapter 2) and is therefore of particular interest. Hsc70-3, the functional counterpart of BiP is thought to play a role in facilitating the assembly of multimeric protein complexes and its upregulation upon Aβ induction is likely to be a protective mechanism against increasing misfolded protein load.[95] The upregulation of BiP is a key component in ER stress responses and this data, therefore, is further evidence towards the expression of Aβ42 inducing ER stress, most likely as a result of a large increase in misfolded protein load.

Another differentially expressed protein which plays a direct role in maintaining protein fold homeostasis is the upregulated ubiquilin, **Figure 3.21C**. Ubiqulin physically associates with the proteasome and ubiquitin ligases and is therefore thought to link the ubiquitination machinery to the proteasome to aid misfolded protein degradation. Ubiqulin has many direct links to misfolding neurodegenerative disease including AD. Overexpression has been shown to promote presenilin accumulation and is known to directly interact with both PS1 and PS2 [96], [97]. Using immunohistochemistry it was found that ubiqulin was prominent within the neurons of human brains

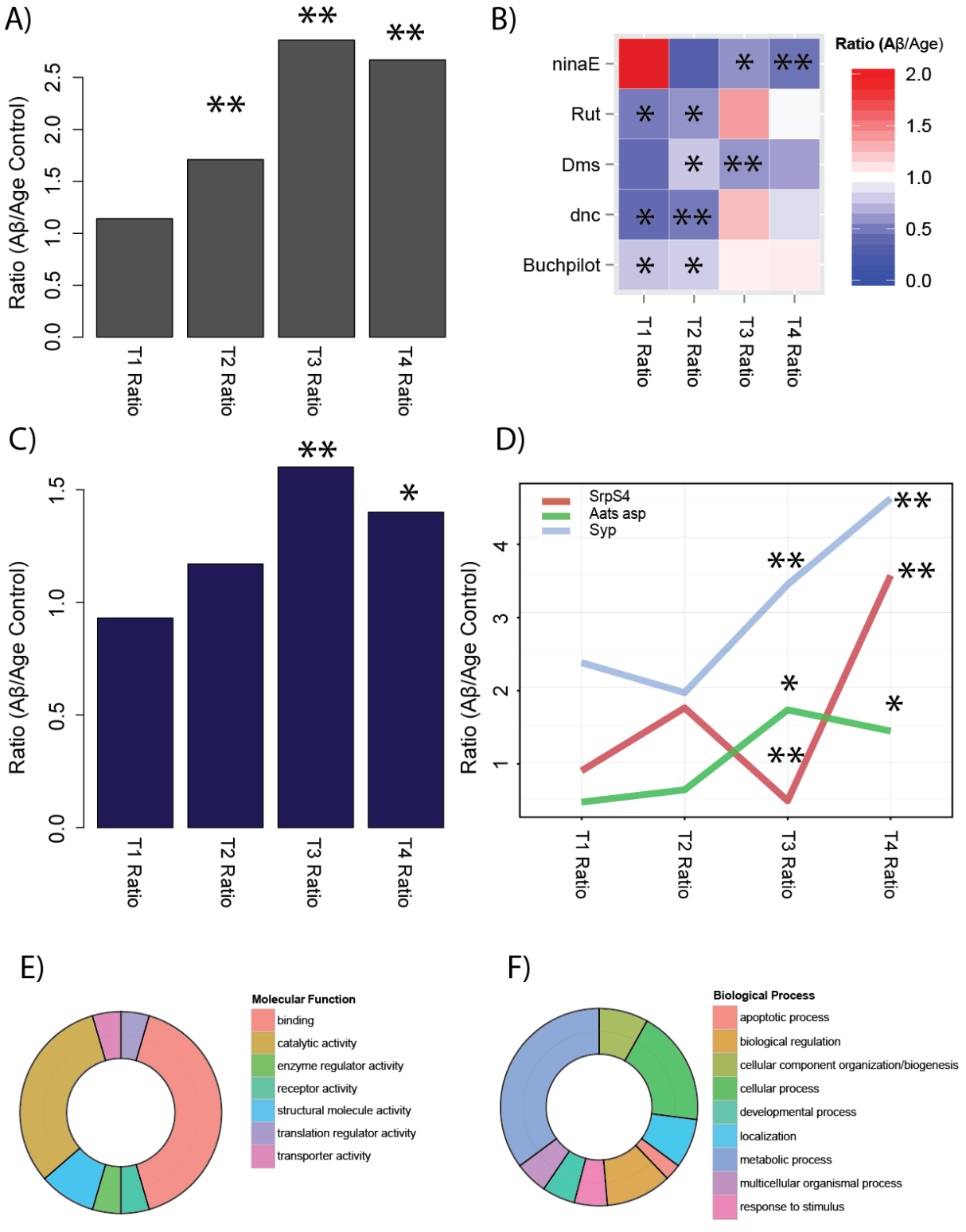


Figure 3.21 Functional analysis of differentially expression proteins as a result of Aβ toxicity in comparison to age matched controls. **A)** Hsc-70 ratios (Aβ/control) across the time-course. **B)** Heatmap of ratios for proteins involved in learning, memory and/or locomotor behaviour. **C)** Ubiquilin ratios across the time-course. **D)** Lineplot of calculated ratios for proteins implicated in neurogenesis. **E&F)** Gene ontology annotation all 44 differentially expressed proteins. * donates p <0.05 **denotes p <0.005

and was also identified to co-localisation with neurofibrillary tangles [96]. Interestingly the overexpression of ubiquilin has been shown to negatively modify the lifespan and motor neuron phenotypes of a *Drosophila* model for ALS [98]. Our findings alongside previous literature provide strong evidence towards ubiquilin being a potentially key modifier of Aβ toxicity through its increasing abundance in respect to age. A small set of proteins implicated in learning, memory and/or locomotor control were identified as differentially expressed and calculated ratios (Aβ/control) are shown in **Figure 3.21B.** It can be seen that all proteins are downregulated in response to Aβ induction and notably this reduction in abundance is first identified, for all 5 proteins, prior to any increase in mortality rates. This is unsurprising as the functions of these proteins are likely to be more closely linked to the negative geotaxis phenotype which occurs weeks prior to an increase in mortality [99]. The identification of bruchpilot is of interest due to its close relation with Rab-3 which was identified as downregulated in the presence of Aβ peptide in the previous study (Chapter 2). Rab-3 has been shown to effect the localisation of bruchpilot at active sites thereby reducing neurotransmitter release [100]. Together the data from the two studies indicate Aβ toxicity to be affecting synaptic dysfunction through the dysregulation of active zone site composition. Three proteins involved in neurogenesis were found to have variable expression levels; however a trend of upregulation at the later time points was observed **Figure 3.21D.** As previously mentioned neurogenesis was until recently not thought to occur in the adult *Drosophila* but recent data is now challenging that assumption [101]. The upregulation of proteins implicated in neurogenesis could, therefore, represent a late compensatory mechanism for Aβ toxicity. Gene ontology analysis of the molecular function and biological process ontologies of all 44 proteins identified a range of different cellular processes but no particular processes dominated the findings. **Figure 3.21 E&F.** A large proportion of proteins identified either had binding or catalytic activity for molecular function characterisation whilst metabolic processes were again seen to be well represented by the data.

3.4 Conclusion

The cellular mechanisms that link ageing to the probability of developing Alzheimer's disease are an important avenue of research to help tackle the socioeconomic challenge of an ageing population. Using a recently developed acquisition method, that uses ion mobility to separate peptide ions prior to ToF mass analysis, HDMS^E, longitudinal quantitative proteomics was undertaken to probe brain proteome alterations as a result of wild-type ageing and chronic Aβ42 misfolding. Approximately 2000 proteins were confidently and consistently quantified across time points than spanned >80% of each of the lifespans from the two conditions. The experimental designed was such that the data were tested in three distinct approaches, allowing protein expression alterations to be successfully correlated with ageing and Aβ toxicity.

Analysis of protein abundances across six time points corresponding to wild-type ageing were interrogated using a soft clustering approach. A coordinated decrease in levels of proteins involved in energy derivation was identified alongside decreased levels of proteins that play a role in signal transduction and sexual reproduction. Potentially decreased levels of energy production and signal transduction could have serious implications for neural plasticity and neural development. Indeed both have been previously linked to neurodegeneration during ageing [54], [74], [102], [103]. Conversely increased abundances for proteins involved in oxidoreducatase activity, endopeptidase activity and catabolic processes were quantified. Data, therefore, is suggestive of increased levels of oxidative stress and degradative processes both have which have been reported to be a cellular consequence of ageing [37], [23], [79]. Following this, a statistical approach that has been optimised for time-course data successfully identified >300 differentially regulated proteins, representing ~15% of all proteins quantified. PPI analysis found >200 of the regulated proteins to directly interact with at least one other protein. A site of high coordination was found to relate to mitochondrial and metabolic proteins, validating the soft clustering results. A second highly coordinated site was found to contain mRNA and ribosomal proteins suggesting translation rates to be a contributing factor in the ageing process. Comparison of our results with data from a quantitative proteomic time-course analysis of ageing in *C.elegans,* recently published by the Mann and Hartl labs, identified conserved ageing processes including reduced ribosomal levels, a dysregulation of the proteosome complex and a large reduction in Sod-3 levels.

A similar approach was used to identify cellular responses to the neuronal expression of an Aβ2 variant with a high aggregating propensity. Soft cluster analysis identified some similarly enriched processes to those found in the ageing data. Again a reduction in mitochondrial and reproductive levels alongside an increase in oxidoreducatase activity was identified. Interestingly proteins involved in actin cytoskeleton organisation were found to reduce in abundance throughout the 4 time points examined. Cytoskeletal abnormalities have long been linked to AD and other neurodegenerative disease [75], [76]. Our data indicates that these processes begin at the very initial stages of AB42 aggregation. Finally, an increase in fatty acid oxidation was observed after T2 as process that has been previously linked to AD [77]. Statistical analysis identified fewer differentially expressed proteins than found from the ageing data. PPI analysis produced a smaller interaction network but four distinct clusters were identified within Cytoscape. Small heat shock proteins (similar to those identified in Chapter 2), cytoskeletal proteins, and proteins involved in neuronal process were all identified and represent interesting targets for follow-up functional studies. Differentially regulated proteins common to both the ageing and Aβ datasets were examined. It was found that expression levels almost exclusively correlated between the two conditions. Interestingly the induction of AB42 appears to amplify the rate of expression changes suggesting that the aggregates may aggravate some of the cellular process that occur during ageing. A number of proteins were successfully validated against a transcriptomic study of Aβ42 expression in *Drosophila.* mRNA and protein levels were found to correlating extremely well across the two studies. These proteins are, therefore, strong candidates for protein expression changes that occur during the pre-clinical and clinical stages of AD. Follow-up functional studies should help determine how they related to the link between ageing and AD prevalence.

The third and final analysis approach was to identify significantly altered protein abundances between the two conditions. Due to the experimental design the approximately 40 proteins identified by this approach can be attributed to be a direct cellular response to Aβ42 expression. Whether these changes are protective or deleterious is unfortunately beyond the scope of these experiments but can be estimated based on prior functional information. Proteins involved in neurogenesis were all upregulated in response to Aβ42 whilst proteins implicated in learning, memory and locomotor behavior were all downregulated. These are likely to directly relate to common phenotypes observed in this AD model [11]. Two proteins of particular interest were identified based on gene ontology

information. Firstly, Hsc70-3, which was also identified in the previous study (Chapter 2) was consistently upregulated in response to Aβ42 expression. Based on the known function of Hsc70-3 its upregulation is likely a protective response to ER stress [90], [95]. Also upregulated was ubiquilin, which has been previously linked to AD through its interaction with PS1 and PS2 and its colonialization with neurofibrillary tangles.[96] Upregulation of ubiquilin has been shown to reduce proteosomal degradation rates and our data identifies it as a potential key modulator of Aβ42 toxicity [97].

Our longitudinal approach has successfully identified cellular alterations in *D. melanogaster* associated with the ageing process, chronic Aβ42 toxicity upon ageing and Aβ42 toxicity alone. Data has identified known regulators of ageing and Aβ42 toxicity but has also implicated previously unknown responses. Collectively our data is a rich and unique resource for future reductionist and functional studies on the link between ageing and AD in *D. melanogaster.*

3.5 References

[1] R. T. Linn, P. A. Wolf, D. L. Bachman, J. E. Knoefel, J. L. Cobb, A. J. Belanger, E. F. Kaplan, and R. B. D'Agostino, "The 'Preclinical Phase' of Probable Alzheimer's Disease A 13-Year Prospective Study of the Framingham Cohort Title," *JAMA Neurol.*, vol. 52, no. 5, pp. 485–490, 1995.

[2] K. E. Pike, G. Savage, V. L. Villemagne, S. Ng, S. A. Moss, P. Maruff, C. A. Mathis, W. E. Klunk, C. L. Masters, and C. C. Rowe, "β-amyloid imaging and memory in non-demented individuals: Evidence for preclinical Alzheimer's disease," *Brain*, vol. 130, no. 11, pp. 2837–2844, 2007.

[3] R. S. Doody, R. G. Thomas, M. Farlow, T. Iwatsubo, B. Vellas, S. Joffe, K. Kieburtz, R. Raman, X. Sun, P. S. Aisen, E. Siemers, H. Liu-Seifert, and R. Mohs, "Phase 3 trials of solanezumab for mild-to-moderate Alzheimer's disease.," *N. Engl. J. Med.*, vol. 370, no. 4, pp. 311–21, 2014.

[4] S. Salloway, R. Sperling, N. C. Fox, K. Blennow, W. Klunk, M. Raskind, M. Sabbagh, L. S. Honig, A. P. Porsteinsson, S. Ferris, M. Reichert, N. Ketter, B. Nejadnik, V. Guenzler, M. Miloslavsky, D. Wang, Y. Lu, J. Lull, I. C. Tudor, E. Liu, M. Grundman, E. Yuen, R. Black, and H. R. Brashear, "Two phase 3 trials of bapineuzumab in mild-to-moderate Alzheimer's disease," *N. Engl. J. Med.*, vol. 370, no. 4, pp. 322–33, 2014.

[5] J. Hardy and D. J. Selkoe, "The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics.," *Science*, vol. 297, no. 5580, pp. 353–356, 2002.

[6] K. Blennow and H. Hampel, "CSF markers for incipient Alzheimer's disease," *Lancet Neurol.*, vol. 2, no. 10, pp. 605–613, 2003.

[7] A. Hye, S. Lynham, M. Thambisetty, M. Causevic, J. Campbell, H. L. Byers, C. Hooper, F. Rijsdijk, S. J. Tabrizi, S. Banner, C. E. Shaw, C. Foy, M. Poppe, N. Archer, G. Hamilton, J. Powell, R. G. Brown, P. Sham, M. Ward, and S. Lovestone, "Proteome-based plasma biomarkers for Alzheimer's disease," *Brain*, vol. 129, no. 11, pp. 3042–3050, 2006.

[8] S. T. Ferreira, M. N. N. Vieira, and F. G. De Felice, "Soluble protein oligomers as emerging toxins in Alzheimer's and other amyloid diseases.," *IUBMB Life*, vol. 59, no. 4–5, pp. 332–345, 2007.

[9] M. E. Larson and S. E. Lesné, "Soluble Aβ oligomer production and toxicity," *J. Neurochem.*, vol. 120, pp. 125–139, 2012.

[10] J. P. Allen, P. Bacteria, R. E. Blankenship, M. T. Madigan, C. E. Bauer, D. Holten, J. Deisenhofer, J. R. Norris, S. Diego, W. Kaiser, R. J. Silbey, H. Sumi, A. Warshel, S. R. Greenfield, M. R. Wasielewski, D. K. Hanson, R. M. Pearlstein, Z. T. Chu, W. W. Parson, M. Marchi, D. Chandler, J. D. Muller, C. A. Wraight, G. U. Nienhaus, G. C. Walker, T. J. Kang, P. F. Barbara, T. Fonseca, G. R. Fleming, D. W. Oxtoby, R. A. Marcus, T. P. Sakmar, F. Van Mourik, G. Van Der Zwan, S. Haacke, and M. Chergui, ". In addition to this, for five of the mutants studied, it was possible to perform an independent check of the DDG," *Biochemistry*, vol. 316, pp. 750–754, 2007.

[11] K. Iijima, H.-P. Liu, A.-S. Chiang, S. A. Hearn, M. Konsolaki, and Y. Zhong, "Dissecting the pathological effects of human Aβ40 and Aβ42 in Drosophila: A potential model for Alzheimer's disease," *Proc. Natl. Acad. Sci.*, vol. 101, no. 17, pp. 6623–6628, 2004.

[12] M. B. Feany and W. W. Bender, "A Drosophila model of Parkinson's disease," *Neurodegeneration*, vol. 404, no. March, 2000.

[13] O. Sofola-Adesakin, J. I. Castillo-Quan, C. Rallis, L. Tain, I. Bjedov, I. Rogers, L. Li, P. Martinez, M. Khericha, M. Cabecinha, J. Bähler, and L. Partridge, "Lithium suppresses Aβ pathology by inhibiting translation in an adult Drosophila model of Alzheimer's disease," *Front. Aging Neurosci.*, vol. 6, no. JUL, pp. 1–10, 2014.

[14] S. Mizielinska, S. Grönke, T. Niccoli, C. E. Ridler, E. L. Clayton, A. Devoy, T. Moens, F. E. Norona, I. O. C. Woollacott, J. Pietrzyk, K. Cleverley, A. J. Nicoll, S. Pickering-brown, J. Dols, M. Cabecinha, O. Hendrich, P. Fratta, E. M. C. Fisher, L. Partridge, and A. M. Isaacs, "C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins," *Science*, vol. 345, no. 6201, pp. 1192–1195, 2014.

[15] C. W. Wittmann, M. F. Wszolek, J. M. Shulman, P. M. Salvaterra, J. Lewis, M. Hutton, and M. B. Feany, "Tauopathy in Drosophila: neurodegeneration without neurofibrillary tangles," *Science*, vol. 293, no. 5530, pp. 711–714, 2001.

[16] O. Sofola, F. Kerr, I. Rogers, R. Killick, H. Augustin, C. Gandy, M. J. Allen, J. Hardy, S. Lovestone, and L. Partridge, "Inhibition of GSK-3 ameliorates Abeta pathology in an adult-onset Drosophila model of Alzheimer's disease.," *PLoS Genet.*, vol. 6, no. 9, 2010.

[17] I. Rogers, F. Kerr, P. Martinez, J. Hardy, S. Lovestone, and L. Partridge, "Ageing increases vulnerability to aβ42 toxicity in Drosophila.," *PLoS One*, vol. 7, no. 7, p. e40569, 2012.

[18] J. R. Wiśniewski, K. Duś-Szachniewicz, P. Ostasiewicz, P. Ziółkowski, D. Rakus, and M. Mann, "Absolute Proteome Analysis of Colorectal Mucosa, Adenoma, and Cancer Reveals Drastic Changes in Fatty Acid Metabolism and Plasma Membrane Transporters," *J. Proteome Res.*, p. 150817095905008, 2015.

[19] H. Zhou, S. Di Palma, C. Preisinger, M. Peng, A. N. Polat, A. J. Heck, and S. Mohammed, "Toward a comprehensive characterization of a human cancer cell phosphoproteome," *J. Proteome Res.*, vol. 12, no. 1, pp. 260–271, 2013.

[20] M. J. Chalmers, C. L. Mackay, C. L. Hendrickson, S. Wittke, M. Waiden, H. Mischak, D. Fliser, I. Just, and A. G. Marshall, "Combined top-down and bottom-up mass spectrometric approach to characterization of biomarkers for renal disease," *Anal. Chem.*, vol. 77, no. 22, pp. 7163–7171, 2005.

[21] S. Lehnert, S. Jesse, W. Rist, P. Steinacker, H. Soininen, S. K. Herukka, H. Tumani, M. Lenter, P. Oeckl, B. Ferger, B. Hengerer, and M. Otto, "ITRAQ and multiple reaction monitoring as proteomic tools for biomarker search in cerebrospinal fluid of patients with Parkinson's disease dementia," *Exp. Neurol.*, vol. 234, no. 2, pp. 499–505, 2012.

[22] M. Basso, S. Giraudo, D. Corpillo, B. Bergamasco, L. Lopiano, and M. Fasano, "Proteome analysis of human substantia nigra in Parkinson's disease," *Proteomics*, vol. 4, no. 12, pp. 3943–3952, 2004.

[23] D. A. Butterfield, M. Perluigi, and R. Sultana, "Oxidative stress in Alzheimer's disease brain: New insights from redox proteomics," *Eur. J. Pharmacol.*, vol. 545, no. 1, pp. 39–50, 2006.

[24] Q. Huang, L. Yang, J. Luo, L. Guo, Z. Wang, X. Yang, W. Jin, Y. Fang, J. Ye, B. Shan, and Y. Zhang, "SWATH enables precise label-free quantification on proteome scale," *Proteomics*, vol. 15, no. 7, pp. 1215–1223, 2015.

[25] S. Beck, A. Michalski, O. Raether, M. Lubeck, S. Kaspar, C. Baessmann, D. Hornburg, F. Meier, I. Paron, and N. A. Kulak, "The impact II , a very high resolution quadrupole proteomics instrument for deep shotgun," *Mol. Cell. Proteomics*, pp. mcp–M114, 2015.

[26] A. L. Richards, A. E. Merrill, and J. J. Coon, "Proteome sequencing goes deep," *Curr. Opin. Chem. Biol.*, vol. 24, pp. 11–17, 2015.

[27] T. Osterwalder, K. S. Yoon, B. H. White, and H. Keshishian, "A conditional tissue-specific transgene expression system using inducible GAL4.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, no. 22, pp. 12596–601, 2001.

[28] J. Cox, M. Y. Hein, C. A. Luber, I. Paron, N. Nagaraj, and M. Mann, "MaxLFQ allows accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction," *Mol. Cell. Proteomics*, pp. M113.031591–, 2014.

[29] J. T. Leek, E. Monsen, A. R. Dabney, and J. D. Storey, "EDGE: Extraction and analysis of differential gene expression," *Bioinformatics*, vol. 22, no. 4, pp. 507–508, 2006.

[30] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.," *Nat. Protoc.*, vol. 4, no. 1, pp. 44–57, 2009.

[31] H. Mi, A. Muruganujan, and P. D. Thomas, "PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees.," *Nucleic Acids Res.*, vol. 41, no. Database issue, pp. D377–86, 2013.

[32] F. M, "Mfuzz: Soft clustering of time series gene expression data," *R Packag. version 2.28.0*, 2015.

[33] A. Franceschini, D. Szklarczyk, S. Frankild, M. Kuhn, M. Simonovic, A. Roth, J. Lin, P. Minguez, P. Bork, C. von Mering, and L. J. Jensen, "STRING v9.1: protein-protein interaction networks, with increased coverage and integration.," *Nucleic Acids Res.*, vol. 41, no. Database issue, pp. D808–15, 2013.

[34] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, "Cytoscape : A Software Environment for Integrated Models of Biomolecular Interaction Networks Cytoscape : A Software Environment for Integrated Models of Biomolecular Interaction Networks," *Genome Res.*, vol. 12, no. 11, pp. 2498–2504, 2003.

[35] J. Cai, G. Chen, and J. Wang, "ClusterViz: a Cytoscape plugin for graph clustering and visualization," *Sch. Inf. Sci. Eng.*, pp. 2–4, 2010.

[36] C. G. Arsene, R. Ohlendorf, W. Burkitt, C. Pritchard, A. Henrion, G. O'Connor, D. M. Bunk, and B. Güttler, "Protein quantification by isotope dilution mass spectrometry of proteolytic fragments: Cleavage rate and accuracy," *Anal. Chem.*, vol. 80, no. 11, pp. 4154–4160, 2008.

[37] T. Morawe, C. Hiebel, A. Kern, and C. Behl, "Protein homeostasis, aging and Alzheimer's disease," *Mol. Neurobiol.*, vol. 46, no. 1, pp. 41–54, 2012.

[38] O. Philipson, A. Lord, M. Lalowski, R. Soliymani, M. Baumann, J. Thyberg, N. Bogdanovic, T. Olofsson, L. O. Tjernberg, M. Ingelsson, L. Lannfelt, H. Kalimo, and L. N. G. Nilsson, "The Arctic amyloid-B precursor protein (ABPP) mutation results in distinct plaques and accumulation of N- and C-truncated Aβ," *Neurobiol. Aging*, vol. 33, no. 5, pp. 1010.e1–1010.e13, 2012.

[39] C. Nilsberth, A. Westlind-Danielsson, C. B. Eckman, M. M. Condron, K. Axelman, C. Forsell, C. Stenh, J. Luthman, D. B. Teplow, S. G. Younkin, J. Näslund, and L. Lannfelt, "The 'Arctic' APP mutation (E693G) causes Alzheimer's disease by enhanced Abeta protofibril formation.," *Nat. Neurosci.*, vol. 4, no. 9, pp. 887–893, 2001.

[40] D. C. Crowther, K. J. Kinghorn, E. Miranda, R. Page, J. A. Curry, F. A. I. Duthie, D. C. Gubb, and D. A. Lomas, "Intraneuronal Abeta, non-amyloid aggregates and neurodegeneration in a Drosophila model of Alzheimer's disease.," *Neuroscience*, vol. 132, no. 1, pp. 123–35, 2005.

[41] Y. Q. Zhang, D. B. Friedman, Z. Wang, E. Woodruff, L. Pan, J. O'donnell, and K. Broadie, "Protein expression profiling of the drosophila fragile X mutant brain reveals up-regulation of monoamine synthesis.," *Mol. Cell. Proteomics*, vol. 4, no. 3, pp. 278–290, 2005.

[42] K. C. Abruzzi, J. Rodriguez, J. S. Menet, J. Desrochers, A. Zadina, W. Luo, S. Tkachev, and M. Rosbash, "Drosophila CLOCK target gene characterization: Implications for circadian tissue-specific gene expression," *Genes Dev.*, vol. 25, no. 22, pp. 2374–2386, 2011.

[43] S. J. Geromanos, C. Hughes, S. Ciavarini, J. P. C. Vissers, and J. I. Langridge, "Using ion purity scores for enhancing quantitative accuracy and precision in complex proteomics samples," *Anal. Bioanal. Chem.*, vol. 404, no. 4, pp. 1127–1139, 2012.

[44] K. C. Hansen, L. Kiemele, O. Maller, J. O'Brien, A. Shankar, J. Fornetti, and P. Schedin, "An in-solution ultrasonication-assisted digestion method for improved extracellular matrix proteome coverage.," *Mol. Cell. Proteomics*, vol. 8, no. 7, pp. 1648–1657, 2009.

[45] T. Glatter, C. Ludwig, E. Ahrné, R. Aebersold, A. J. R. Heck, and A. Schmidt, "Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion," *J. Proteome Res.*, vol. 11, no. 11, pp. 5145–5156, 2012.

[46] U. Distler, J. Kuharev, P. Navarro, Y. Levin, H. Schild, and S. Tenzer, "Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics.," *Nat. Methods*, vol. 11, no. 2, pp. 167–70, 2014.

[47] N. J. Bond, P. V Shliaha, K. S. Lilley, and L. Gatto, "Improving Qualitative and Quantitative Performance for MSE-based Label-free Proteomics," *J. Proteome Res*, vol.12, no. 6, pp. 2340-2353, 2013.

[48] J. Cox, M. Y. Hein, C. A. Luber, I. Paron, N. Nagaraj, and M. Mann, "Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ," *Mol. Cell. Proteomics*, vol. 13, pp. 2513–2526, 2014.

[49] G.-Z. Li, J. P. C. Vissers, J. C. Silva, D. Golick, M. V Gorenstein, and S. J. Geromanos, "Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures.," *Proteomics*, vol. 9, no. 6, pp. 1696–719, Mar. 2009.

[50] Y. Levin, E. Hradetzky, and S. Bahn, "Quantification of proteins using data independent analysis (MS(E)) in simple and complex samples: a systematic evaluation.," *Proteomics*, pp. 3273–3287, Jun. 2011.

[51] R. C. Brown, A. H. Lockwood, and B. R. Sonawane, "Neurodegenerative diseases: An overview of environmental risk factors," *Environ. Health Perspect.*, vol. 113, no. 9, pp. 1250–1256, 2005.

[52] H. F. Poon, R. A. Vaishnav, T. V. Getchell, M. L. Getchell, and D. A. Butterfield, "Quantitative proteomics analysis of differential protein expression and oxidative modification of specific proteins in the brains of old mice," *Neurobiol. Aging*, vol. 27, no. 7, pp. 1010–1019, 2006.

[53] D. M. Walther and M. Mann, "Accurate quantification of more than 4000 mouse tissue proteins reveals minimal proteome changes during aging.," *Mol. Cell. Proteomics*, vol. 10, no. 2, p. M110.004523, 2011.

[54] A. Trifunovic and N. G. Larsson, "Mitochondrial dysfunction as a cause of ageing," *J. Intern. Med.*, vol. 263, no. 2, pp. 167–178, 2008.

[55] M. K. Shigenaga, T. M. Hagen, and B. N. Ames, "Oxidative damage and mitochondrial decay in aging.," *Proc. Natl. Acad. Sci.*, vol. 91, no. 23, pp. 10771–10778, 1994.

[56] D. M. Walther, P. Kasturi, M. Zheng, S. Pinkert, G. Vecchi, P. Ciryam, R. I. Morimoto, C. M. Dobson, M. Vendruscolo, M. Mann, and F. U. Hartl, "Widespread Proteome Remodeling and Aggregation in Aging C. elegans," *Cell*, vol. 161, no. 4, pp. 919–932, 2015.

[57] K. G. Iliadi, A. Avivi, N. N. Iliadi, D. Knight, A. B. Korol, E. Nevo, P. Taylor, M. F. Moran, N. G. Kamyshev, and G. L. Boulianne, "nemy encodes a cytochrome b561 that is required for Drosophila learning and memory.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 50, pp. 19986–19991, 2008.

[58] C. T. Chien, S. Wang, M. Rothenberg, L. Y. Jan, and Y. N. Jan, "Numb-associated kinase interacts with the phosphotyrosine binding domain of Numb and antagonizes the function of Numb in vivo.," *Mol. Cell. Biol.*, vol. 18, no. 1, pp. 598–607, 1998.

[59] E. J. Masoro and S. N. Austad, "The evolution of the antiaging action of dietary restriction: a hypothesis.," *J. Gerontol. A. Biol. Sci. Med. Sci.*, vol. 51, no. 6, pp. B387–B391, 1996.

[60] J. N. Keller, F. F. Huang, and W. R. Markesbery, "Decreased levels of proteasome activity and proteasome expression in aging spinal cord," *Neuroscience*, vol. 98, no. 1, pp. 149–156, 2000.

[61] R. Roubenoff, "Catabolism of aging: is it an inflammatory process?," *Curr. Opin. Clin. Nutr. Metabloic Care*, vol. 6, no. 3, pp. 295–299, 2003.

[62] S. M. Esslinger, B. Schwalb, S. Helfer, K. M. Michalik, H. Witte, K. C. Maier, D. Martin, B. Michalke, A. Tresch, P. Cramer, and K. Förstemann, "Drosophila miR-277 controls branched-chain amino acid catabolism and affects lifespan.," *RNA Biol.*, vol. 10, no. 6, pp. 1042–56, 2013.

[63] A. Terman and U. T. Brunk, "Aging as a catabolic malfunction," *Int. J. Biochem. Cell Biol.*, vol. 36, no. 12, pp. 2365–2375, 2004.

[64] K. Hirai, G. Aliev, A. Nunomura, H. Fujioka, R. L. Russell, C. S. Atwood, A. B. Johnson, Y. Kress, H. V Vinters, M. Tabaton, S. Shimohama, A. D. Cash, S. L. Siedlak, P. L. Harris, P. K. Jones, R. B. Petersen, G. Perry, and M. A. Smith, "Mitochondrial abnormalities in Alzheimer's disease.," *J. Neurosci.*, vol. 21, no. 9, pp. 3017–3023, 2001.

[65] J. W. Lustbader, M. Cirilli, C. Lin, H. W. Xu, K. Takuma, N. Wang, C. Caspersen, X. Chen, S. Pollak, M. Chaney, F. Trinchese, S. Liu, F. Gunn-Moore, L.-F. Lue, D. G. Walker, P. Kuppusamy, Z. L. Zewier, O. Arancio, D. Stern, S. S. Yan, and H. Wu, "ABAD directly links Abeta to mitochondrial toxicity in Alzheimer's disease.," *Science*, vol. 304, no. 5669, pp. 448–452, 2004.

[66] M. T. Lin and M. F. Beal, "Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases.," *Nature*, vol. 443, no. 7113, pp. 787–795, 2006.

[67] I. Bjedov, J. M. Toivonen, F. Kerr, C. Slack, J. Jacobson, A. Foley, and L. Partridge, "Mechanisms of Life Span Extension by Rapamycin in the Fruit Fly Drosophila melanogaster," *Cell Metab.*, vol. 11, no. 1, pp. 35–46, 2010.

[68] K. Jia, D. Chen, and D. L. Riddle, "The TOR pathway interacts with the insulin signaling pathway to regulate C. elegans larval development, metabolism and life span.," *Development*, vol. 131, no. 16, pp. 3897–3906, 2004.

[69] G. V. De Ferrari and N. C. Inestrosa, "Wnt signaling function in Alzheimer's disease," vol. 33, pp. 1–12, 2000.

[70] J. Höhfeld, Y. Minami, and F. U. Hartl, "Hip, a novel cochaperone involved in the eukaryotic Hsc70/Hsp40 reaction cycle.," *Cell*, vol. 83, no. 4, pp. 589–598, 1995.

[71] C. T. Murphy, S. A. McCarroll, C. I. Bargmann, A. Fraser, R. S. Kamath, J. Ahringer, H. Li, and C. Kenyon, "Genes that act downstream of DAF-16 to influence the lifespan of Caenorhabditis elegans.," *Nature*, vol. 424, no. 6946, pp. 277–283, 2003.

[72] A. Hsu, C. T. Murphy, and C. Kenyon, "Regulation of Aging and Age-Related Disease by DAF-16 and Heat-Shock Factor," vol. 936, no. June, 2003.

[73] D. M. Kurnit, R. L. Neve, L. S. Disease, A. D. Is, M. L. V. A. N. Keuren, D. Patterson, and S. Pagan, "Amyloid beta protein gene: cDNA, mRNA distribution, and genetic linkage near the Alzheimer locus," *Science*, vol. 235, no. 4791, pp. 880–884, 1987.

[74] J. L. Chou, D. V Shenoy, N. Thomas, P. K. Choudhary, F. M. Laferla, S. R. Goodman, and G. A. M. Breen, "Early dysregulation of the mitochondrial proteome in a mouse model of Alzheimer's disease.," *J. Proteomics*, vol. 74, no. 4, pp. 466–79, 2011.

[75] G. Perry, M. Kawai, M. Tabaton, M. Onorato, P. Mulvihill, P. Richey, A. Morandi, J. A. Connolly, and P. Gambetti, "Neuropil threads of Alzheimer's disease show a marked alteration of the normal cytoskeleton.," *J. Neurosci.*, vol. 11, no. 6, pp. 1748–1755, 1991.

[76] J. P. Brion, "The pathology of the neuronal cytoskeleton in Alzheimer's disease," *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.*, vol. 1160, no. 1, pp. 134–142, 1992.

[77] T. J. Montine and J. D. Morrow, "Fatty acid oxidation in the pathogenesis of Alzheimer's disease.," *Am. J. Pathol.*, vol. 166, no. 5, pp. 1283–1289, 2005.

[78] G. M. Cole, G. P. Lim, F. Yang, B. Teter, A. Begum, Q. Ma, M. E. Harris-White, and S. A. Frautschy, "Prevention of Alzheimer's disease: Omega-3 fatty acid and phenolic anti-oxidant interventions," *Neurobiol. Aging*, vol. 26, no. SUPPL., pp. 133–136, 2005.

[79] W. R. Markesbery, "Oxidative Stress Hypothesis in Alzheimer's Disease," *Free Radic. Biol. Med.*, vol. 23, no. 1, pp. 134–147, 1997.

[80] R. Kaddurah-Daouk, H. Zhu, S. Sharma, M. Bogdanov, S. G. Rozen, W. Matson, N. O. Oki, A. A. Motsinger-Reif, E. Churchill, Z. Lei, D. Appleby, M. A. Kling, J. Q. Trojanowski, P. M. Doraiswamy, and S. E. Arnold, "Alterations in metabolic pathways and networks in Alzheimer's disease.," *Transl. Psychiatry*, vol. 3, no. 4, p. e244, 2013.

[81] R. Kaddurah-Daouk, S. Rozen, W. Matson, X. Han, C. M. Hulette, J. R. Burke, P. M. Doraiswamy, and K. A. Welsh-Bohmer, "Metabolomic changes in autopsy-confirmed Alzheimer's disease," *Alzheimer's Dement.*, vol. 7, no. 3, pp. 309–317, 2011.

[82] K. Kamino, K. Nagasaka, M. Imagawa, H. Yamamoto, H. Yoneda, A. Ueki, S. Kitamura, K. Namekata, T. Miki, and S. Ohta, "Deficiency in mitochondrial aldehyde dehydrogenase increases the risk for late-onset Alzheimer's disease in the Japanese population.," *Biochem. Biophys. Res. Commun.*, vol. 273, no. 1, pp. 192–196, 2000.

[83] C. S. Ferguson and R. F. Tyndale, "Cytochrome P450 enzymes in the brain: Emerging evidence of biological significance," *Trends Pharmacol. Sci.*, vol. 32, no. 12, pp. 708–714, 2011.

[84] A. Elbaz and C. Levecque, "CYP2D6 Polymorphism , Pesticide Exposure , and Parkinson ' s Disease," *Ann. Neurol.*, vol. 55, no. 3, pp. 430–434, 2004.

[85] I. P. Nezis, A. Simonsen, A. P. Sagona, K. Finley, S. Gaumer, D. Contamine, T. E. Rusten, H. Stenmark, and A. Brech, "Ref(2)P, the Drosophila melanogaster homologue of mammalian p62, is required for the formation of protein aggregates in adult brain," *J. Cell Biol.*, vol. 180, no. 6, pp. 1065–1071, 2008.

[86] S. P. Koushika, M. J. Lisbin, and K. White, "ELAV, a Drosophila neuron-specific protein, mediates the generation of an alternatively spliced neural protein isoform.," *Curr. Biol.*, vol. 6, no. 12, pp. 1634–1641, 1996.

[87] B. Charroux, C. Angelats, L. Fasano, S. Kerridge, and C. Vola, "The levels of the bancal product, a Drosophila homologue of vertebrate hnRNP K protein, affect cell proliferation and apoptosis in imaginal disc cells," *Mol. Cell. Biol.*, vol. 19, no. 11, pp. 7846–7856, 1999.

[88] V. Dwivedi, B. K. Tripathi, M. Mutsuddi, and S. C. Lakhotia, "Ayurvedic amalaki rasayana and rasa-Sindoor suppress neurodegeneration in fly models of Huntington's and Alzheimer's diseases," *Curr. Sci.*, vol. 105, no. 12, pp. 1711–1723, 2013.

[89] J. Moreno, H. Radford, D. Peretti, Steinert, N. Verity, M. Martin, M. Halliday, J. Morgan, D. Dinsdale, C. Ortori, D. Barrett, P. Tsaytler, A. Bertolotti, A. Willis, M. Bushell, and G. Mallucci, "Sustained translational repression by eIF2α-P mediates prion neurodegeneration.," *Nature*, vol. 485, no. 7399, pp. 507–511, 2012.

[90] T. Katayama, K. Imaizumi, T. Manabe, J. Hitomi, T. Kudo, and M. Tohyama, "Induction of neuronal death by ER stress in Alzheimer's disease," *J. Chem. Neuroanat.*, vol. 28, no. 1–2, pp. 67–78, 2004.

[91] S. J. Marciniak and D. Ron, "Endoplasmic Reticulum Stress Signaling in Disease," *Physiol Rev*, vol. 86, no. 4, pp. 1133–1149, 2006.

[92] G. Favrin, D. M. Bean, E. Bilsland, H. Boyer, B. E. Fischer, S. Russell, D. C. Crowther, H. A. Baylis, S. G. Oliver, and M. E. Giannakou, "Identification of novel modifiers of Aβ toxicity by transcriptomic analysis in the fruitfly.," *Sci. Rep.*, vol. 3, p. 3512, 2013.

[93] D. G. Ballinger, N. Xue, and K. D. Harshman, "A Drosophila photoreceptor cell-specific protein, calphotin, binds calcium and contains a leucine zipper.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 90, no. 4, pp. 1536–1540, 1993.

[94] A. Federico, E. Cardaioli, P. Da Pozzo, P. Formichi, G. N. Gallus, and E. Radi, "Mitochondria, oxidative stress and neurodegeneration," *J. Neurol. Sci.*, vol. 322, no. 1–2, pp. 254–262, 2012.

[95] Y. Kimata, Y. I. Kimata, Y. Shimizu, H. Abe, I. C. Farcasanu, M. Takeuchi, M. D. Rose, and K. Kohno, "Genetic Evidence for a Role of BiP / Kar2 That Regulates Ire1 in Response to Accumulation of Unfolded Proteins," *Mol. Biol. Cell*, vol. 14, no. June, pp. 2559–2569, 2003.

[96] A. L. Mah, G. Perry, M. A. Smith, and M. J. Monteiro, "Identification of ubiquilin, a novel presenilin interactor that increases presenilin protein accumulation," *J. Cell Biol.*, vol. 151, no. 4, pp. 847–862, 2000.

[97] L. Massey, A. Mah, D. Ford, J. Miller, J. Liang, H. Doong, and M. Monteiro, "Overexpression of ubiquilin decreases ubiquitination and degradation of presenilin proteins," *J. Alzheimer's Dis.*, vol. 6, no. 1, pp. 79–92, 2004.

[98] K. A. Hanson, S. H. Kim, D. A. Wassarman, and R. S. Tibbetts, "Ubiquilin modifies TDP-43 toxicity in a Drosophila model of amyotrophic lateral sclerosis (ALS)," *J. Biol. Chem.*, vol. 285, no. 15, pp. 11068–11072, 2010.

[99] F. Kerr, H. Augustin, M. D. W. Piper, C. Gandy, M. J. Allen, S. Lovestone, and L. Partridge, "Dietary restriction delays aging, but not neuronal dysfunction, in Drosophila models of Alzheimer's disease.," *Neurobiol. Aging*, vol. 32, no. 11, pp. 1977–89, Nov. 2011.

[100] E. R. Graf, R. W. Daniels, R. W. Burgess, T. L. Schwarz, and A. DiAntonio, "Rab3 dynamically controls protein composition at active zones.," *Neuron*, vol. 64, no. 5, pp. 663–77, Dec. 2009.

[101] I. Fernández-Hernández, C. Rhiner, and E. Moreno, "Adult Neurogenesis in Drosophila," *Cell Rep.*, vol. 3, no. 6, pp. 1857–1865, 2013.

[102] G. S. Roth and J. A. Joseph, "Cellular and molecular mechanisms of impaired dopaminergic function during aging," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 719, no. 13, pp. 129–35, 1994.

[103] L. Jin and T. Saitoh, "Changes in Protein Kinases in Brain Aging and Alzheimer's Disease," *Drugs and Ageing*, vol. 6, no. 2, pp. 136–149, 1995.

Subcellular Proteomics of α_1-Antitrypsin Misfolding in a Mammalian Cell Model of ER Stress and Disease

4.0 Abstract

Folding of secreted proteins within the endoplasmic reticulum (ER) requires interaction of ER quality control mechanisms with the folding polypeptide chain. The metastable antiprotease α_1-antitrypsin is synthesised in large amounts in the liver and is one of the most highly secreted human proteins. The $Null_{HongKong}$ (NHK) variant has a major C-terminal deletion and so grossly misfolds in the ER. The proteostatic challenge generated by its expression makes this an important model for ER stress and triggering of the unfolded protein response. We have used label-free (MS^E), ER-focused proteomics in the CHO-K1 mammalian cell model, to define the molecular response to folding and secretion of high levels of wild-type (M) protein. These were compared with the ER-associated proteome obtained with expression of the NHK variant to identify subcellular responses to α_1-antitrypsin misfolding. The data indicate changes affecting regulation of proteostatic load, protein folding and protein processing within the ER. Fascinatingly our studies also indicate altered bioenergetics and mitochondrial behaviour. Data derived from live cell fluorescence microscopy validate these changes by demonstrating that misfolding within the ER induces dramatic changes in mitochondrial morphology between cells expressing M, NHK and the common severe disease variant Z α_1-antitrypsin. The changes show genotype:phenotype correlation and indicate that the cellular response to α_1-antitrypsin misfolding within the ER may include modulation of the balance between mitochondrial fusion and fission.

4.1 Introduction

Discovered in 1965 by Laurell and Erikssoon, the folding of the archetypal serpin (serine protease inhibitor) α_1-antitrypsin within the hepatocyte endoplasmic reticulum (ER) has been observed to challenge human proteostatic quality control mechanism [1]. α_1-antitrypsin's main physiological function is to inhibit neutrophil elastase but also functions to inhibit two other proteases, also from the neutrophil family, cathepsin G and proteinase 3. Circulating levels of α_1-antitrypsin carefully maintain homeostatic levels of neutrophil elastase to facilitate the removal of damaged or ageing lung tissue cells whilst preventing the uncontrolled breakdown of healthy lung tissue.

Common with other serpins it displays a distinctive suicidal mechanism for inhibiting proteases. A result of this is that the 45 kDa polypeptide chain adopts a complex native fold. Unlike most proteins, the native conformation of α_1-antitrypsin does not represent the polypeptide's most stable globular fold. Instead, for it to function as an enzyme inhibitor it must remain kinetically trapped in a metastable conformation [2]. The protein is small and relatively polar, allowing the movement into tissue fluid such as the lungs. In its metastable conformation, as shown in **Figure 4.1**, the protein presents an exposed reactive loop, whose centre is focused around Met358-Ser359. This reactive centre loop acts a substrate for a neutrophil protease. After docking of the target protein, the centre of the target loop is cleaved, leaving the loop more conformationally flexible. Increased flexibility leads to the insertion of the loop as an addition strand within the center of β-sheet A. This insertion forms the inhibitory complex that renders both the protease and serpin inactive, following a conformational translocation of the protease. The opening of the β-sheet A alongside the insertion of the reactive loop leads to local and global stabilization interactions resulting in a highly stable complex.

α_1-antitrypsin is synthesised by and secreted from hepatocytes in large quantities (1-2 g/d) and as such is one of the most abundant proteins in human serum (20-48 μM). The protein is of particular interest due to the monogenic disorder α_1-antitrypsin deficiency (AATD), a disease which clinically is found to be relatively common in European populations. AATD is linked to chronic liver disease in childhood and hepatocellular carcinoma/cirrhosis in adulthood [3]. The main clinical symptom is however early onset pulmonary emphysema.

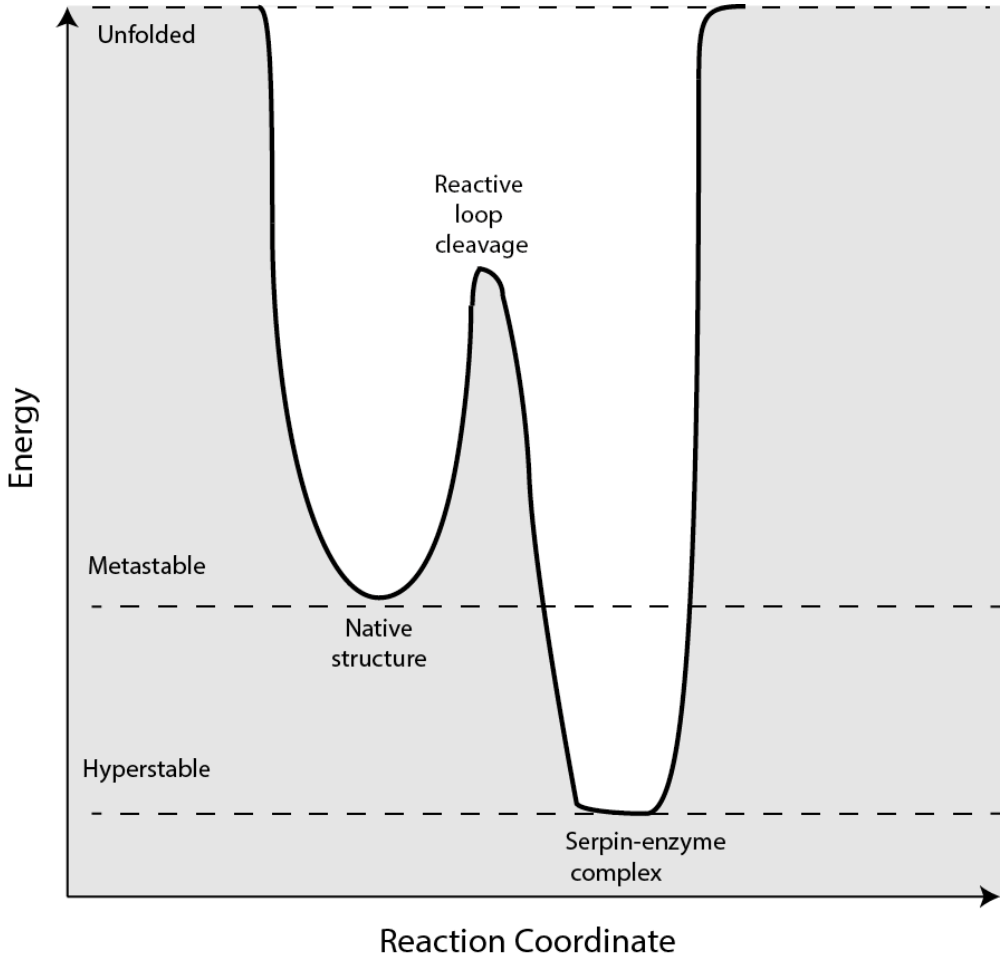


Figure 4.1 Energy diagram illustrating the different conformation states of α1-antitrypsin during enzyme inhibition

Circulating levels of α_1-antitrypsin lower than 11 μM are insufficient to properly inhibit neutrophil elastase. This allows the progressive destruction of alveoli and if left untreated can lead to emphysema. The pathophysiology of AATD in relation to liver disease is far less well understood but it is known that deficient variants can accumulate in the endoplasmic reticulum due to inefficient secretion. This accumulation results in protein aggregation of α_1-antitrypsin variants and causes subsequent hepatocyte damage. AATD is caused by point mutations within the SERPINA1 gene, that encodes for α_1-antitrypsin. More than 100 alleles have been identified at the time of writing; however we will focus on three of the more common variants.

PiM type alleles account for more than 95 % of Caucasian individuals. Individuals that are homozygotes for M alleles produce normal plasma levels (> 20 μM). The $Null_{HongKong}$ (NHK) variant results from a frameshift mutation affecting residue 318 that introduces 15 new residues before

introducing a premature stop codon after position 333, giving rise to a 61 amino-acid C-terminal truncation.[4] These changes cause it to grossly misfold, saturating the ER-associated degradation (ERAD) response with resultant ER stress. Indeed NHK α_1-antitrypsin expression is widely used as an experimental trigger in studies characterising these responses [4]–[9]. ERAD involves the recognition of misfolded protein substrates and their retrotranslocation from the ER into the cytosol for ubiquitination and degradation by proteasomes. ER degradation-enhancing α-mannosidase (EDEM) selectively recognizes misfolded protein by immature *N*-linked glycosylation, and targets these for ERAD. The mechanism underlying protein retrotranslocation is poorly defined, but may involve the action of the Sec61 complex, Derlin-1 and the HRD1-SEL1L complex.[9]–[11] Further complexity is suggested by the finding that NHK α_1-antitrypsin can be cycled via the Golgi, and that mannose modifying and non-modifying roles of mannosidase proteins may also contribute to the overall ERAD process [12], [13].

Saturation of the ERAD pathway causes the ER unfolded protein load to rise, initiating the unfolded protein response (UPR) that generally reduces the rate of polypeptide synthesis and increases ER chaperone capacity [14]. This is triggered by activation of the ER stress sensors inositol-requiring IRE1, double-stranded RNA–activated protein kinase-like ER kinase (PERK), and activating transcription factor ATF6. When UPR and ERAD together fail to adapt the cell to its ER synthetic load, uncontrolled ER stress can result in apoptosis [15].

The PiZ allele is the most common cause of AATD and is associated with very low serum concentrations of protein. Homozygotes for the Z variants possess circulating levels of α_1-antitrypsin in the range of 5-6 µM. It has been shown that 85 % of Z homozygous children have liver function test results outside the expected range, however only 10 % will present clinical symptoms whilst only 1-3 % will develop cirrhosis during childhood [16]. The Z variant is a result of a point mutation of Glu342Lys. Northern Europeans have a high incidence of this mutation with 1 in 2000 being homozygous and 1 in 27 a heterozygote carrier [17]. In Z homozygotes 15 % of α_1-antitrypsin that is synthesized is successfully trafficked to the Golgi for secretion. Retained protein forms self-associated polymer chains and eventually results in hepatocyte inclusion bodies.

Unlike the NHK variant, retention of the Z variant within the ER does not appear to initially activate the UPR, when studied using cells that recapitulate hepatocyte phenotypes [18]. In fact, further to the 15

% of secreted protein roughly 70 % is degraded in the ER with the remaining protein forming polymer aggregates. It is found, however, that cells are instead sensitized to UPR activation upon an additional insult. Serpin polymer accumulation in the ER can trigger an alternative stress response, the ordered protein response or more commonly the ER overload response. Interestingly studies suggest that accumulation of the Z variant activates this response to a far greater degree than the NHK form. The ER overload response is poorly annotated at the present time but studies suggest it to be calcium-dependent and a downstream result is the activation of the transcription factor NF-kB [19].

Quantitative proteomics has been previously used to good effect to study the response of cellular systems to the assault of protein aggregation. A large majority of these studies have recorded the global cellular response, in order to define a disease phenotype. These data however, do not allow the separation of responses from differing cellular compartments. This is of particular importance when the aggregating protein, in our case α_1-antitrypsin is known to localise in a subcellular manner; namely the endoplasmic reticulum.

Sub-cellular fractionation, a commonly employed technique in cellular biology, coupled to mass spectrometry-based proteomics allow the enrichment of cellular components to better define a subcellular proteomic response. The Lilley group have used their technique, localization of organelle proteins by isotope tagging (LOPIT) to good success to better annotate the proteomes of sub cellular compartments [20]. As with much subcellular proteomics it uses a density gradient, whereby organelles are separated following centrifugation across a sucrose or iodixanol gradient. Organelles can be successfully separated based on their different physical properties. This preparation technique allows the direct use of cellular homogenate, however in all cases does not produce single organelle fractions as some compartments share similar properties. In reality a distribution of proteins from overlapping organelles is produced, with different organelles enriched in different fractions. The Lilley group have further developed this technique with the addition of multivariate data analysis to better define the overlapping organelle protein distributions. Label based quantification, in the form of iTRAQ is employed to measure protein amounts across the density gradient. Quantification of proteins with distinct organelle associations can be used to create a phenotypic representation of organelle specific protein distribution. These distributions can then be used to assign proteins of unknown association.

The use of antibodies to detect known residents of a subcellular compartment is still a widely used technique for isolating a specific organelle for subsequent interrogation [21], [22]. A known marker allows the normalisation of repeat fractionation protocols, to ensure that the extracted proteome is consistent across repeat biological runs. This approach benefits MS-based proteomic approaches, which are often limited by instrumental dynamic range. The removal of organelles of disinterest facilitates the identification of a larger number of proteins of interest, which may have otherwise been unidentified, due to co-eluting tryptic peptides from high abundance proteins originating from removed organelles. What must be considered with this approach is however, the possibility of co-fractionating proteins from organelles distinct from the compartment of interest. Indeed it is known that mitochondrial proteins commonly co-elute with resident endoplasmic reticulum proteins during density gradient centrifugation [23]. Functional analysis post MS analysis can be used to clarify the localisation of many of the identified proteins, but will not provide complete coverage.

Table 4.1. Characterized ER focused cellular responses to the expression of α_1-antitrypsin variants

Cellular Response	M (Wild Type)	$_{null}$Hong Kong	Z
Secretion	+++	-	+/- (~15%)
ER associated degradation (ERAD)	+	++	++
Autophagy	+/-	+/-	+ to +/-
Unfolded protein response (UPR)	-	+++	-
ER overload response (EOR)	-	-	++

We wished to define the subcellular proteomic responses to folding and secretion of the metastable wild-type, to misfolding-induced degradation of NHK α_1-antitrypsin and Z α_1-antitrypsin. To date the responses to α_1-antitrypsin misfolding have been principally characterised by classical reductionist

cell biology methods. An overview of these expected responses is highlighted in **Table 4.1**, responses it was hoped could be further expanded upon by our methodology. In our highly complementary approach we therefore undertook label-free (MS^E), unbiased, ER-focused proteomic studies in a non-malignant, mammalian cell expression system (CHO-K1, *Tet-On*) for secreted proteins that fold within the ER. This epithelial-like line is also validated as a good phenotypic model for the expression of α_1-antitrypsin variants in hepatocytes [24]. We obtained 6 ER-focused proteomic profiles associated with low (non-induced) and high (doxycline-induced) expression of wild-type, NHK α_1-antitrypsin and Z α_1-antitrypsin. The approach is enriched for response components of high relevance to the secretory pathway and misfolding responses that will localise to the ER itself.

4.2 Materials and Methods

4.2.1 Preparation of CHO Cell ER Samples

Chinese hamster ovary (CHO)-K1 cells stably transfected with Tet-On expression systems for M , $Null_{HongKong}$ (NHK) or Z variants of α_1-antitrypsin were cultured, and expression induced with doxycycline, as described previously.[24] Non-induced cell lines were not treated with doxycycline but otherwise cultured identically, **Figure 4.2A**. Cells were harvested after 5 days seeded (2 days post-induction), and lysed by 25-30 passages through an EMBL cell homogenizer (Isobiotec, Heidelberg, Germany) loaded with an 8.010 mm ball bearing. The procedure was carried out on ice, and cell disruption was confirmed by light microscopy.

Homogenized samples were centrifuged (3000 rpm, Eppendorf 5415D centrifuge, rotor F45-24-11, 5-10 minutes) to pellet out cell nuclei. Density gradient centrifugation was used to purify the ER fractions using an OptiPrep (Axis-Shield, Oslo, Norway) iodixanol gradient system. 600 µl of each sample made up to 2 ml of 35% (w/v) iodixanol solution was transferred to the bottom of a centrifuge tube. An Auto Densi-Flow Density Gradient Fractionator (Labconco, Kansas City) was used to create a density gradient of 30% to 10% (w/v) iodixanol on top of the sample layer, and samples were then centrifuged for 2 hours at 100,000*g* (23,500 rpm using an SW40i rotor, Beckman Coulter, CA, USA) at 4 °C. 0.5 ml fractions were collected sequentially from the top of the tube. Samples from each fraction were run on SDS-PAGE and assessed by western blot analysis for the presence of α_1-antitrypsin and resident ER chaperones (Grp-98, BiP and protein disulphide isomerase (PDI)) containing the KDEL motif, **Figure 4.2B**. ER-associated fractions (defined as all KDEL-positive fractions) were then pooled for further analysis and concentrated to <200 µl. All steps involved in CHO ER sample preparation were carried out by Dr Peg Nyon, Gooptu Lab, UCL.

4.2.2 GeLC-MS^E Analysis

Pooled samples were resolved by 1D SDS-PAGE and stained with Coomassie Blue. Each lane was divided into seven sections, **Figure 4.2C**, which were then treated to undergo tryptic digestion as described previously [25]. Briefly, bands were destained, disulphide bonds were reduced with dithiothreitol and free cysteine residues were alkylated with iodoacetamide. Gel bands were washed

and dehydrated prior to the addition of a 10 ng/µl tryptic solution (Promega, Southampton, UK) in 50 mM ammonium bicarbonate. Digestion was carried out overnight at 37 °C. Resultant tryptic peptides were extracted in a two-step process. Solutions were then lyophilised to dryness and stored at -20 °C prior to analysis.

Before liquid chromatography (LC)-MS^E analysis, digested samples were solubilised in 10 µl of 0.1% formic acid. For each sample 1 µl was loaded using a split-less nano-Ultra Performance Liquid Chromatography (10 kpsi nanoAcquity Waters Corporation, Milford, MA, USA). Buffers used were (A) water + 0.1% formic acid (J.T. Baker, PA, USA) and (B) acetonitrile + 0.1% formic acid (J.T. Baker, PA, USA). Samples were desalted using a reverse-phase 2G V/M C18 trap column (180 µm internal diameter, 20 mm length, 5 µm particle size, Waters Corporation) at a flow rate of 5 µl/min for 3 minutes. Peptides were separated by a three step gradient (0.4 µl/min, 55 °C; 99-92% Buffer A over 7.5 mins, 92-66% Buffer A 62.5 mins, 66-60% 5 mins) using a HSS T3 C18 nano-column (75 µm internal diameter, 150 mm length, 1.7 µm particle size, Waters Corporation).

The nanoLC was coupled online through a nanoESI emitter (7 cm length, 10 mm tip; New Objective, Woburn, MA, USA) to a quadrupole time-of-flight (QToF) hybrid mass spectrometer (HDMS Synapt; Waters UK) tuned to a mass resolution of 10 000 (full width half height). The ToF analyzer was externally calibrated from *m/z* 175.11 to 1285.54 using the fragment ions of a 500 fmol/µl solution of [Glu^1]-fibrinopeptide B (GFP, Sigma Aldrich, St.Louis, MO). Data were lockmass-corrected post acquisition using the monoisotopic mass of the doubly charged precursor of GFP (785.8426 *m/z*), delivered at 500 fmol/µl to the mass spectrometer via a NanoLockSpray interface.

Mass measurements were acquired in MS^E positive mode, whereby the quadrupole mass analyser is set to transmit all ions. The collision energy within the transfer region is alternated from low to high throughout analysis facilitating the acquisition of both MS and MS/MS for all ions entering the mass spectrometer. The radio-frequency applied to the quadrupole was set such that ion transmission from *m/z* 300-2000 was maximized. This ensured that any ions observed below *m/z* 300 could be confidently assigned as fragment ions from dissociations within the collision cell. For low-energy (MS) mode a constant 4 eV was applied while an energy ramp from 15 to 35 eV was used for high-energy

mode (MS/MS). Scan time was set to 1 s with a 0.05 s interscan delay and the reference sprayer was sampled every 60 s. 2 technical replicate datasets were obtained for each of 3 biological replicate experiments for each cell line.

4.2.3 MS^E Data Processing

LC-MS^E data were processed using ProteinLynx Global Server (PLGS) v2.5.2. The data processing prior to database searching included ion detection, alignment and clustering methods detailed elsewhere [26]. Briefly, lockmass corrected spectra were centroided, deisotoped and charge state reduced. Initial assignment of fragment ions to parent peptides was achieved by retention time-alignment. Thresholds used for peak selection were as follows, 150 counts for low energy ion detection, 40 counts for high energy ion detection and 500 counts for precursor exact mass retention time (EMRT) integrated intensity.

4.2.4 Database Searches

Data were searched using PLGS v2.5.2 against a Chinese Hamster Ovary (CHO) database (www.chogenome.org/files/CHO_refseq_protein.fasta) appended with human α_1-antitrypsin and common contaminant proteins. Carbamidomethyl (C) was set as a fixed modification while oxidation (M), acetylation (K & N-term) and deamidation (N & Q) were set as variable modifications. Precursor and fragment ion tolerances were determined automatically by PLGS. Protein identification criteria were set as the detection of at least 2 fragments ion per peptide, at least 2 fragments per protein and at least 1 peptide identification per protein. A maximum of 2 missed cleavages were allowed. The protein level false discovery rate was maintained at < 5%, estimated based upon the number of proteins identified from a randomized database. Post-processing the false discovery rate was minimized further by applying the additional criterion that for a successful identification a protein must be identified in a minimum of two biological repeats (see Chapter 2.2.6).The mass spectrometry data and associated database search results have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD002219 and 10.6019/PXD002219.

4.2.5 Protein Quantification

Protein quantification was achieved through the Hi3 technique. Hi3 is based upon the observation that the summed intensity of the three most ionizing peptides per protein correlates with high accuracy and precision to protein concentration [27]. As label-free quantification directly compares samples analysed in concurrent runs quantification errors can be introduced by run-to-run instrument variability. To reduce this within-condition normalization was applied prior to quantification. This was undertaken at the peptide level using a script written in-house in R to implement locally-weighted scatter point smoothing, **Figure 4.2D**. Post LOESS normalization, further global normalization was applied based upon the total detected protein intensity per run, in order to account for sample preparation and loading variability across biological replicates and across conditions. Differential α_1-antitrypsin levels resulted in only minimal change to total protein levels and so are unlikely to have had an appreciable effect on the normalisation process. The ratio of α_1-antitrypsin levels to total protein levels were: 0.0%, 0.01%, 0.02%, and 0.12% respectively for CHO transfected cell lines with M and NHK α_1-antitrypsin expression non-induced and for M and NHK α_1-antitrypsin expression induced. Based upon total protein intensity it was found that biological replicates from the same condition gave good reproducibility, **Figure 4.2D**, with 22% variability in average total protein levels. Comparison of non-induced to induced conditions showed a similar trend with a 25% difference between NHK+ and NHK- and only a 2% difference between M- and M+. The largest difference in protein amounts was found between the two induced samples, M+ and NHK+, where the analysis of M+ samples consistently identified around 70% more protein than the analysis of NHK+ samples. This is consistent with ER stress conditions where protein synthesis is globally reduced. The normalisation steps undertaken were therefore important in our conservative approach to identify robust changes in expression levels that were likely to be of more specific relevance to misfolding and reduced secretion of α_1-antitrypsin.

4.2.6 Differential Expression

Pairwise comparisons of relative expression levels were performed for the six conditions (M and NHK α_1-antitrypsin stable CHO-K1 Tet-On cell lines, without (-) or with (+) α_1-antitrypsin expression induced by doxycycline treatment). Differential expression was calculated based on fold change and statistical analysis. Fold change was calculated for each comparison using the mean of Hi3 protein intensity

across replicates. P values were calculated using a two-sided Student's t-test, with equal variance and was performed on $\log_2$ transformed intensity data. Proteins that were calculated to have a fold change larger than 1.3-fold and a p value < 0.05 were defined as differentially expressed. MS^E quantitative technical variation has been previously demonstrated to be reproducible at 10-15%. A 1.3-fold threshold therefore represents a value typically 2-3 times larger than estimated technical measurement error and has previously been shown to compare well with more established label-based methods [28]. Statistical results are provides in the appendix file.

An additional category in the comparison of ER-associated proteomes defined proteins as unique between cell lines if they were confidently identified in one condition (present in ≥2 biological repeats) but was not detected at all in the other (see Chapter 2.2.10). Lastly a subset of proteins were detected in ≥2 biological repeats in one cell line, and in 1 biological repeat in the other. Such proteins fell between criteria for "unique" or "differentially expressed" proteins in common between cell lines. They were therefore not taken forward for further analysis, but are provided in the appendix file.

4.2.7 Data Storage, LOESS Normalisation and Extraction

Database search results, as Final fragment.csv ion accounts, were exported from PLGS and loaded into a custom built MySQL database that stores data in a hierarchical structure, thus maintaining the experimental structure. Peptide intensities derived from the same protein and identified in adjacent gel bands were summed prior to normalisation. Within-condition normalisation was conducted on summed peptide intensities, using a R-based script employing the loess {stats} function. Briefly, between-sample correlation was calculated and the sample with the highest average correlation was taken forward as the sample to which all were normalised. MA plots (log peptide intensity ratios vs mean peptide intensity) were calculated for all samples and a local regression line fitted. A weighting factor was calculated for each residual and applied to the dataset such that post normalisation the local regression fit is close to M= 0A. Data were extracted from the database using custom queries for post processing in R. Database construction was carried out by David Sutherland, Thalassinos lab, UCL.

4.2.8 Live Cell Microscopic Visualisation of Mitochondria

To assess mitochondrial morphology, cultured CHO-K1 cells expressing M, NHK, or the common pathogenic Z (Glu342Lys) variants of α_1-antitrypsin mutation were incubated with the MitoTracker Red stain M7512 Red CMXRos (Life Technologies, ThermoFisher, MA, USA) as per manufacturer's instructions. Live cells were imaged on an incubated LSM710 laser point scanning confocal microscope (Carl Zeiss Ltd, Cambridge UK) and acquired with the Zen 2010 software using a 63x 1.4NA Plan Apochromat objective lens. The experiments were repeated 3 times for each cell line. For quantitative analysis of mitochondrial morphology, mitochondria were selected in an automated manner and their sphericity scores calculated using Volocity (PerkinElmer, MA, USA) and Imaris (Bitplane, Zurich, Switzerland) software. All imaging was carried out by Charis-Patricia Segeritz, Cambridge Stem Cell Institute, Cambridge Institute for Medical Research.

4.2.9 Unsupervised Hierarchical Clustering

Pairwise comparisons between non-induced and induced samples of the same variant and between induced samples across variant were carried out to identify regulated proteins across the conditions. Proteins were taken forward for clustering if they were identified in all six conditions, by > 1 biological repeat per condition and were identified as regulated ($p < 0.05$). Average intensities were calculated and $\log_2$ transformed prior to clustering. Unsupervised hierarchical clustering was achieved using the R function heatmap.2 {gplots}. The clustering method used was complete linkage.

4.2.10 BLAST Alignment and Conversion

CHO cell protein accessions were converted to human accession for downstream analysis by the Basic Local Alignment Search Tool (BLAST). An expect threshold of 10 was used as a threshold for acceptance of alignment. The median expect value for all proteins included in the search was $1x10^{-105}$. Where more than one human protein was successfully aligned to a CHO protein, the alignment with the lower expect value was taken forward for analysis.

4.2.11 Enrichment Analysis

Enrichment analysis was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID, http://david.abcc.ncifcrf.gov/) [29]. Proteins identified in the non-induced M variant α_1-antitrypsin CHO cell line, post BLAST conversion were searched against a human proteome background to identify enriched biological process terms.

4.2.12 STRING and Cytoscape Analysis

Following conversion of accession information to human paralogues by BLAST search, proteins identified as regulated by >1.3 fold were submitted into the STRING database search tool (http://string-db.org/) to identify interaction networks [30]. Proteins upregulated were searched separately to those found to be downregulated. All protein-protein interaction prediction methods were included in the search results. Protein-protein interactions with a confidence value of ≥ 0.7 (high confidence) were imported into cytoscape for visualisation [31]. Nodes were arranged by optimisation of the organic layout as provided within cytoscape. Clusters were defined using the ClusterViz application within which the EAGLE algorithm was applied [32].

4.3 Results

4.3.1 Technical Overview

CHO-K1 *Tet-On* cells that were stably transfected to synthesise M or NHK or Z α_1-antitrypsin when induced by doxycycline treatment were assessed in the non-induced (M-, NHK- and Z-) and induced (M+, NHK+ and Z+) states, **Figure 4.2A**. Successful fractionation of the ER was indicated by clustering of the KDEL containing chaperones, **Figure 4.2B**. Moreover, the proteins that were identified were enriched for ER-associations relative to the probabilities of identifying them in a whole cell lysate **Table 4.2**. Highlighted are biological processes which present enrichment ($p < 0.01$ B-H corrected). A large majority of the process agree with the principle functions that occur within the ER, including protein translation, protein folding and protein transport, all which display high level of enrichment.

The MS^E protocol showed good technical and biological reproducibility of quantitation following normalisation across data sets with R^2 correlation values in log-log intensity plots of 0.75-0.85 and 0.7 respectively, **Figure 4.2D**. The good biological reproducibility confirmed the validity of our fraction selection strategy. Whilst small shifts were apparent in the distributions of KDEL-positive fractions between biological replicates, these were very similar between all cell lines when all repeats were considered. Overall no systematic effects were therefore expected to influence our findings that are based on data incorporating findings from all repeats. On the other hand, strict pooling of identically numbered fractions in all repeats would have risked missing a proportion of ER-containing fractions on some occasions and varying enrichment of the pool for ER material overall. 878 proteins were identified for M-, 828 for M+, 884 for NHK-, 941 for NHK+, 719 for Z- and 684 for Z+, **Figure 4.2E**. After filtering based on replication, the false discovery rate (FDR) at the protein level was reduced to < 1% and the peptide level FDR was found to be an average of 0.13%.

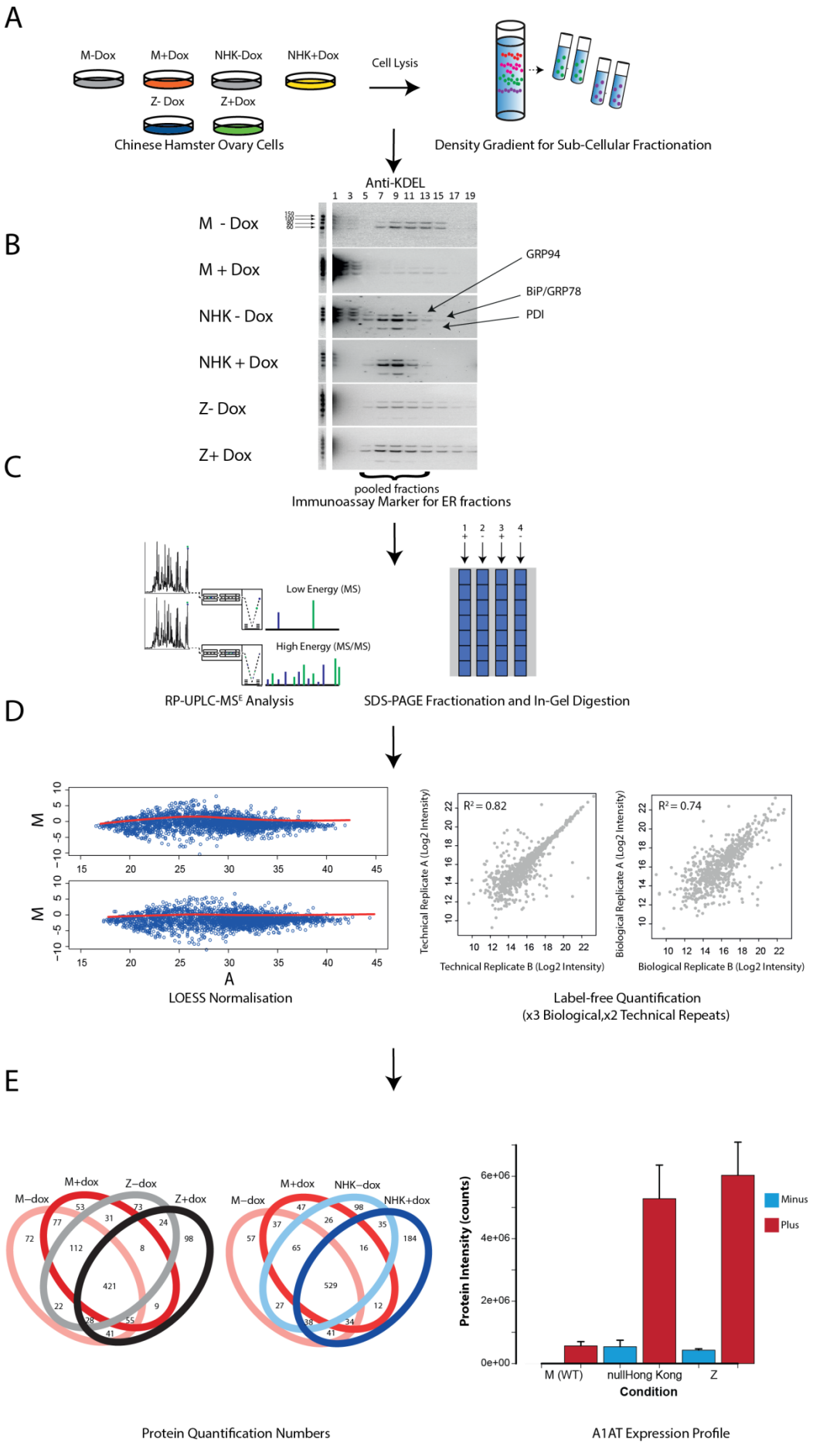

A
M-Dox
M+Dox
NHK-Dox
NHK+Dox
Z- Dox
Z+Dox
Chinese Hamster Ovary Cells
Cell Lysis
Density Gradient for Sub-Cellular Fractionation
B
Anti-KDEL
1 3 5 7 9 11 13 15 17 19
M - Dox
M + Dox
NHK - Dox
NHK + Dox
Z- Dox
Z+ Dox
GRP94
BiP/GRP78
PDI
C
pooled fractions
Immunoassay Marker for ER fractions
Low Energy (MS)
High Energy (MS/MS)
RP-UPLC-MS^E Analysis
SDS-PAGE Fractionation and In-Gel Digestion
D
M
A
LOESS Normalisation
$R^2 = 0.82$
Technical Replicate A (Log2 Intensity)
Technical Replicate B (Log2 Intensity)
$R^2 = 0.74$
Biological Replicate A (Log2 Intensity)
Biological Replicate B (Log2 Intensity)
Label-free Quantification
(x3 Biological,x2 Technical Repeats)
E
M–dox
M+dox
Z–dox
Z+dox
NHK–dox
NHK+dox
Protein Quantification Numbers
Protein Intensity (counts)
Minus
Plus
M (WT)
null
Hong Kong
Z
Condition
A1AT Expression Profile

Figure 4.2 Subcellular quantitative proteomic characterisation of ER response to α_1-antitrypsin misfolding. A) Expression of M, NHK and Z α_1-antitrypsin variants was induced with doxycycline in CHO-K1 cells transfected with *Tet-On* expression systems. Following cell lysis, density gradient centrifugation was used to separate organelles B) ER specific fractions were selected on the basis of SDS-PAGE/western blot positivity for ER resident chaperones containing KDEL motifs. Fraction numbers are shown at the top, migration of marker ladder indicated (left) with masses indicated in kDa, α_1-antitrypsin transfection and induction states are labelled (right). Bracket below lowest western blot indicates fractions pooled for that fractionation. C) Pooled fractions were separated by SDS-PAGE and digested with trypsin. The tryptic digest peptides were then separated by UPLC prior to analysis by the "data-independent" acquisition mass spectrometry method, MS^E. D) To correct for instrumental variability between runs, LOESS normalisation was applied at the peptide level within conditions (left). Here M represents the calculated ratio between the same peptide, identified in different samples and plotted on a log2 scale ($\log_2\{P_1/P_2\}$). A denotes half the log2 value of peptide summed intensity $1/2\log_2(P_1P_2)$. Post-normalisation R^2 values for technical repeats were 0.8-0.99 and 0.7 for biological repeats, based on log-log plots (right). E) Left panel, qualitative comparisons between cell lines in terms of their ER-associated proteomes. Proteins identified as common between 2 cell lines were used for quantitative comparisons Right panel, α_1-antitrypsin n expression levels were successfully quantified in induced samples (M+, NHK+ and Z+) and agreed with expected levels from known cellular processing. Background leakiness was observed in the NHK- and Z-samples. Error bars shown are standard error of the mean (SEM).

Relative quantitation was conducted on an identified protein if it was detected with a minimum of three peptides. Quantitative data post-normalisation is provided in the appendix file. Post-filtering, 87.6% of the proteins in the data set were quantifiable by these criteria. This high proportion reflects the benefits of using the more comprehensive approach afforded by the power of MS^E analysis (data-independent approach) in comparison to a conventional workflow (where selected peptides are analysed using a data-dependent approach). Peptides that were unambiguously assignable to proteins were preferentially selected for quantification purposes.

4.3.2 Analysis of Proteomic Profiles

Qualitatively, differences in the studied ER-associated proteomes were indicated by different profiles of KDEL-containing chaperones observed on western blot analysis of the cellular fractions **Figure 4.2B**. This suggested that total levels of protein disulphide isomerase (PDI) were increased when

NHK α_1-antitrypsin was expressed, even at low levels **Figure 4.2E**, relative to the wild-type M protein. The effects were less pronounced when comparing Z variant expression to wild-type demonstrating a phenotypic difference between the two samples. Proteomic profiling allowed far more exhaustive and quantitative assessment. Subcellular α_1-antitrypsin profiles were consistent with the known cellular handling of the three variants. Thus in the M+ system, minimal α_1-antitrypsin was observed within the cell whilst expression of NHK α_1-antitrypsin showed a signal of intracellular retention within the ER and cytoplasm **Figure 4.2E** and **Figure 4.3**. Although not statistically significant the Z+ system was found to have further increased α_1-antitrypsin levels which can be attributed to the retention of self-ordered polymers that are not present in the NHK+ system. These findings are consistent with misfolded NHK and Z α_1-antitrypsin recognition by the ERAD machinery and retrotranslocation into the cytoplasm. Background leakiness of the expression system was indicated by MS detection of NHK and Z α_1-antitrypsin retention in the ER in the absence of doxycycline induction. It is likely that M α_1-antitrypsin was produced at similar levels in the non-induced state, but not detected due to its highly efficient secretion from the cell. These data illustrate the superior sensitivity of the MS methods applied in this study relative to western blot analyses, which were unable to detect background α_1-antitrypsin levels present in NHK- cells, **Figure 4.3**.

Comparison of data from the different cell lines allowed assessment of their similarity based upon the relative abundance of proteins identified in all 6 lines using hierarchical clustering **Figure 4.4A**. It was found that all non-induced samples clustered together, indicating low levels of cellular response in the presence of low α_1-antitrypsin expression levels. Samples where the two misfolding variants were expressed, NHK & Z, were found to cluster together which is consistent with the variants activating common ER specific responses as a result of their misfolding. Interestingly the system expressing wild-type, M α_1-antitrypsin showed the greatest divergence from all other samples. This is unsurprising as these samples represent basal levels of ER homeostatic mechanisms would not be expected to be as active in non-induced cells but will be highly regulated upon the expression of misfolding variants.

Table 4.2 Results from enrichment analysis of proteins identified by the LC-MS^E analysis of the CHO ER-associated proteome. Proteins queried were those detected in M- cells.

Biological Process Term	Number of Genes	Percentage of Total	P-Value	FDR
translation	165	4.6	8.20E-33	2.30E-29
translational elongation	74	2.1	4.00E-29	5.60E-26
cellular respiration	59	1.7	2.60E-17	2.40E-14
cellular protein metabolic process	634	17.8	1.60E-16	7.60E-14
small GTPase mediated signal transduction	121	3.4	2.40E-14	1.30E-11
protein transport	240	6.7	2.80E-13	1.30E-10
respiratory electron transport chain	40	1.1	2.10E-12	8.20E-10
negative regulation of protein modification process	58	1.6	2.10E-11	7.10E-09
nucleobase, nucleoside, nucleotide and nucleic acid biosynthetic process	81	2.3	2.50E-11	7.70E-09
nucleobase, nucleoside and nucleotide biosynthetic process	81	2.3	2.50E-11	7.70E-09
ribonucleotide biosynthetic process	59	1.7	4.60E-11	1.30E-08
nucleoside phosphate metabolic process	108	3	1.30E-10	3.30E-08
negative regulation of cellular protein metabolic process	75	2.1	2.20E-10	5.10E-08
ATP synthesis coupled electron transport	34	1	3.70E-10	7.80E-08
positive regulation of ligase activity	40	1.1	5.40E-10	1.10E-07
negative regulation of protein metabolic process	76	2.1	6.50E-10	1.20E-07
protein folding	73	2	6.80E-10	1.20E-07
positive regulation of ubiquitin-protein ligase activity during mitotic cell cycle	38	1.1	7.50E-10	1.20E-07
cellular protein complex assembly	68	1.9	1.20E-09	1.80E-07
nucleoside triphosphate biosynthetic process	49	1.4	1.70E-09	2.40E-07
purine ribonucleotide metabolic process	60	1.7	2.50E-09	3.50E-07
negative regulation of ubiquitin-protein ligase activity during mitotic cell cycle	36	1	3.10E-09	4.10E-07
carboxylic acid metabolic process	172	4.8	4.30E-09	5.40E-07
mitochondrial electron transport, NADH to ubiquinone	27	0.8	6.40E-09	7.60E-07
regulation of ubiquitin-protein ligase activity	40	1.1	6.50E-09	7.50E-07
hexose metabolic process	75	2.1	6.70E-09	7.40E-07
purine nucleoside triphosphate metabolic process	54	1.5	8.20E-09	8.70E-07
negative regulation of ligase activity	36	1	8.90E-09	9.10E-07
cellular carbohydrate catabolic process	42	1.2	1.00E-08	1.00E-06
glucose metabolic process	63	1.8	1.20E-08	1.20E-06
acetyl-CoA metabolic process	22	0.6	1.80E-08	1.70E-06
hexose catabolic process	36	1	2.40E-08	2.10E-06
purine nucleotide biosynthetic process	60	1.7	5.50E-08	4.80E-06
monosaccharide catabolic process	36	1	6.10E-08	5.10E-06
mitochondrial transport	35	1	9.60E-08	7.80E-06
proton transport	32	0.9	1.40E-07	1.10E-05
intracellular protein transport	119	3.3	2.60E-07	2.00E-05
coenzyme catabolic process	18	0.5	9.20E-07	6.80E-05
Golgi vesicle transport	52	1.5	1.00E-06	7.40E-05
energy coupled proton transport, down electrochemical gradient	23	0.6	1.70E-06	1.20E-04
ATP synthesis coupled proton transport	23	0.6	1.70E-06	1.20E-04
ion transmembrane transport	26	0.7	2.10E-06	1.40E-04
phosphorylation	215	6	8.40E-06	5.70E-04
cytoskeleton-dependent intracellular transport	25	0.7	3.00E-05	1.90E-03
protein targeting to mitochondrion	19	0.5	3.00E-05	1.90E-03
nucleobase biosynthetic process	11	0.3	5.90E-05	3.70E-03
post-Golgi vesicle-mediated transport	26	0.7	8.30E-05	5.10E-03
protein polymerization	23	0.6	1.10E-04	6.30E-03
RNA processing	149	4.2	1.20E-04	7.00E-03
cell redox homeostasis	27	0.8	1.40E-04	8.30E-03
protein targeting	68	1.9	1.50E-04	8.20E-03
regulation of programmed cell death	209	5.9	1.90E-04	1.00E-02

Cluster analysis therefore successfully validated our approach where we focused our further analyses upon differences observed between M+, NHK+ and Z+ states to characterise responses to the misfolding variant relative to wild-type α_1-antitrypsin that is efficiently secreted. Proteins were identified as unique to one or the other cell type, or common to both. The relative abundances of the latter group of proteins were then compared between the two states in terms of fold-change, and the statistical significance of the difference calculated. To determine cellular response that were unique to the two misfolding variants Z+ and NHK+ were directly compared using the same approach, **Figure 4.4B** .

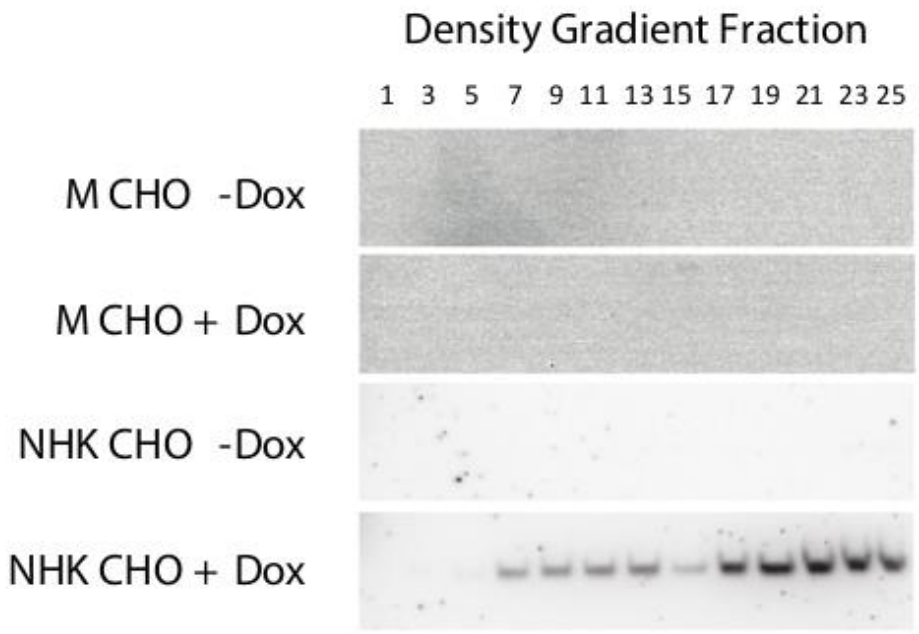


Figure 4.3 Western blot analyses of cell lysate fractions, probing for α_1-antitrypsin in cells transfected with M or NHK α_1-antitrypsin. Cells were cultured in the absence (M-, NHK-) or presence (M+, NHK+) of doxycycline to compare non-induced and induced states with respect to α_1-antitrypsin expression

Some of the proteins identified in this way are database-annotated as ER-resident, whilst others have clear potential for some physical association with ER proteins, based upon their assigned functions. In the remainder, such associations could not be established from current annotation. This last category likely contains some proteins that localise to the ER in a manner that remains to be characterised and/or annotated in the protein database. The other proteins are likely to have co-fractionated for other reasons e.g. general upregulation of that protein within the cell, causing its increased representation across a wide range of cellular fractions. Analysis of the datasets using current annotation does not allow discrimination between these different options concerning apparently 'non-ER' proteins co-fractionating with the ER at this time. However subcellular proteomics offers the ability to provide data on a scale not currently possible by other techniques and so it is an ideal workflow to provide enhanced cellular location annotation. Work by the Lilley lab is therefore highly complementary in this regard, since it combines fractionation of cell contents with multivariate data analysis to group proteins within compartments using the Localisation of Organelle Proteins by Isotope Tagging (LOPIT) approach [20]. For this study we were ultimately interested in understanding

the cellular response to α_1-antitrypsin misfolding, albeit with a focus on upstream ER responses. We therefore elected to analyse change in the 'ER-associated' proteome as a whole. This approach allows ready application of future advances in understanding of protein behaviour to our published dataset. Moreover, it provided the potential to identify strong signals of other intracellular responses to misfolding for further study.

Based on binary comparisons, **Figure 4.4B,** the expression of misfolding variants was observed to induce a number of statistically significant expression changes. Annotation of these expression changes will improve current knowledge of ER specific regulation of protein folding. When the two variants, NHK+ & Z+ are directly compared little difference was observed with few proteins being identified as being statistically significant. Our data, appears to identifying similar ER responses upon the expression of the two misfolding variants. Our work, therefore, focuses on the differences observed between the M+ and NHK+ cell lines. This is based on the results from the binary comparisons suggesting that the majority of our findings will be directly applicable to the Z+ cell lines.

Proteins showing significant differences in abundance were assessed by two independent approaches. Network analysis related these proteins on the basis of previously described interactions between them, **Figure 4.5, Figure 4.7** and **Appendix F**, while pathway analysis was based upon KEGG database annotation of each individual protein within different physiological pathways, **Figure 4.6** and **Appendix G**.

STRING network analysis, **Figure 4.5** and **Appendix F**, of the ER-associated proteome in cells expressing NHK compared with those expressing M α_1-antitrypsin was conducted using stringent criteria. Data were then imported to Cytoscape for further analysis. Misfolding of NHK α_1-antitrypsin induced increased representation of proteins interacting within six principal clusters, defined by the ClusterViz algorithm as implemented within Cytoscape, **Figure 4.5** [31], [32]. Most of these clusters are directly relatable to control of protein load and misfolding within the ER. Cluster I (labelled Gold, **Appendix F**) comprises proteins

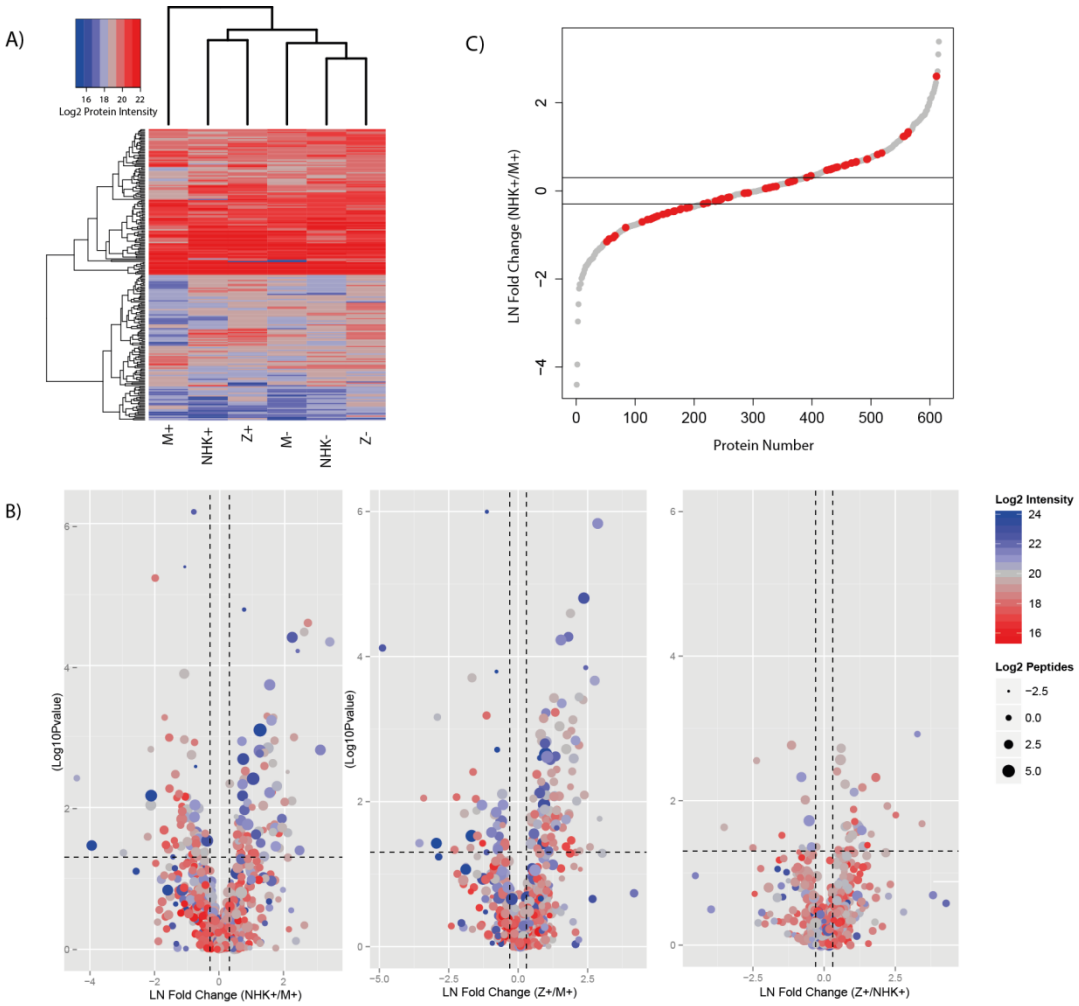


Figure 4.4 Quantitative pairwise comparisons of M+, NHK+ and Z+ expression levels. **A)**Unsupervised hierarchical clustering of regulated proteins, based on pairwise comparisons between all conditions. **B)** Volcano plot of M+ vs NHK+, Z+ vs M+ and Z+ vs NHK+. Plot colours are representative of average Hi3 protein intensity for the most highly expressed protein from each binary comparison. Plot sizes are representative of the average number of unique peptides identified for each protein. Averages are calculated across the conditions and include redundancy from peptides identified in more than one gel band. Proteins whose expression data are plotted in the two upper outer quadrants indicated by the dotted lines are defined as differentially expressed. **C)** Log_2 fold-change for all proteins found in both M+ and NHK+ cells. Proteins defined as ER-specific by gene ontology annotation are highlighted in red

capable of interacting with pre-mRNAs, and/or effects upon mRNA splicing, turnover and transport. These include serine/arginine-rich splicing factor (SRSF)7, small nuclear ribonucleoproteins (SNRPs) D1, D3 and E and heterogeneous nuclear ribonucleoproteins (HNRNPs) C, D, F, R, U & M. It also contains three heat shock proteins (HSPs, A8, A5 and A4) which act as repressors of transcriptional activity. These findings are consistent with the general principle that the ER stress response involves

modification of upstream elements controlling differential splicing and the rate of polypeptide synthesis for folding within the ER [23].

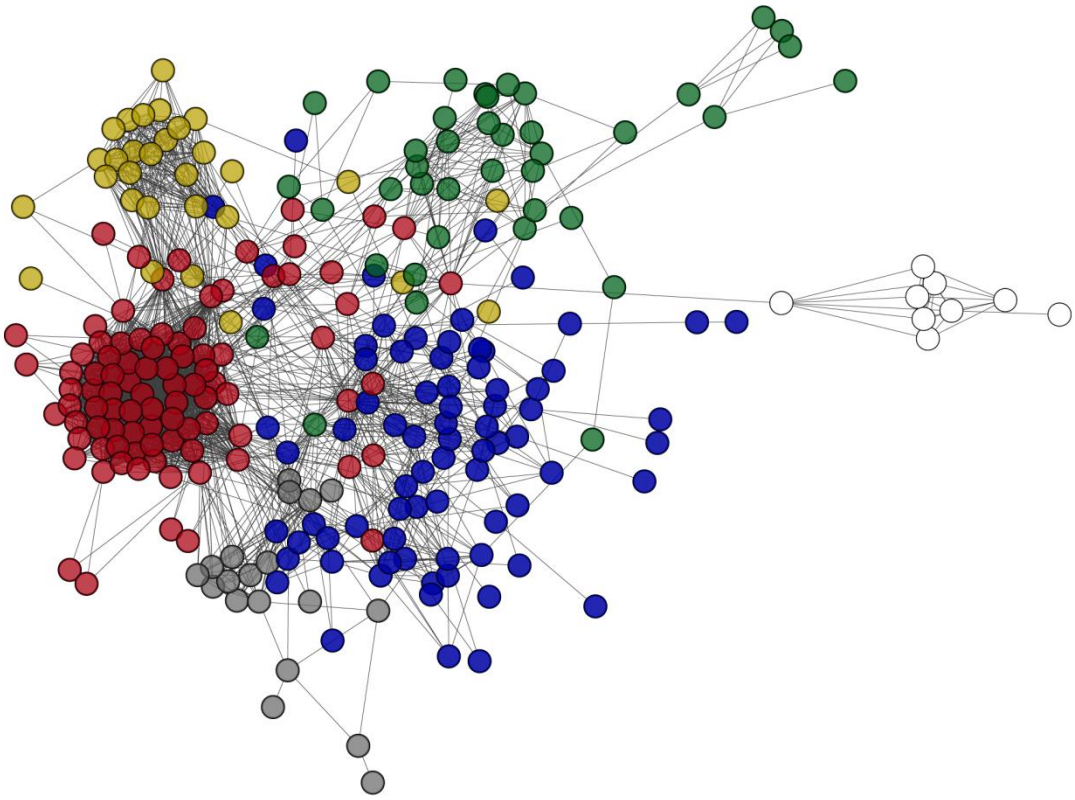


Figure 4.5 Protein-protein interaction network. Protein-protein interaction network (STRING) analysis for proteins with detected expression levels >1.3 fold higher in NHK+ (relative to M+). Only interactions with a score of ≥ 0.7 are included. Interaction network layout is the yFiles organic setting as implemented within cytoscape. Colours denote clusters as calculated by the EAGLE clustering algorithm (ClusterViz plugin) [22]. I, Gold: mRNA handling. II, Red: Ribosomal subunits and translation initiation factors. III, Blue: Proteostasis and cytoskeleton. IV, White: Glutathione processing. V, Green: Bioenergetics. VI, Grey: Ubiquitin, Proteasome and α_1-antitrypsin

Similarly, Cluster II (labeled Red, **Appendix F**) consists of predominantly ribosomal proteins and translation initiation (and elongation) factors (eIFs) such as subunits of eIF2, 3, 4 and 5. It also includes GNB2L1 that can act as a repressor of translation. The same cluster also includes tumor protein, translationally-controlled (TPT)1 that plays a role in phenotypic reprogramming. Cluster III (Blue, **Appendix F**) represents proteins involved in chaperoning and degradation and cytoskeletal elements. The directly proteostatic elements include: valosin-containing protein (VCP)/p97, a ubiquitin segregase known to regulate both proteasome- and lysosome-mediated degradation of substrates from diverse sources [24]; 26*S* proteasome non-ATPase regulatory subunit 2 (PSMD2); the UPR

chaperone heat shock 70 protein 4; and prolyl isomerase FK506 binding protein (FKBP)4, a member of the immunophilin family involved in protein chaperoning and trafficking. It also contains actins and tubulins, raising the possibility that ER stress is coupled to increased interactions with cytoskeleton components. Cluster IV (White, **Appendix F**) consists of proteins involved in glutathione transferase-mediated processes of disulphide bond manipulation in protein folding quality control including glutathione S-transferase P and microsomal glutathione S-transferase 1. Class V (Green, **Appendix F**) relates to bioenergetic pathway elements. These include proteins involved in glycolytic metabolism (phosphoglycerate kinase (PGK), pyruvate kinase M (PKM), gamma enolase (ENO2) and lactate dehydrogenase (LDH) A and B). Class VI (Grey, **Appendix F**) contains α_1-antitrypsin and further proteostatic components, the proteasomal subunits PSMA 1,7,3,8 and PSMB 1,3,4,5, the ubiquitin-related protein USP5.

Pathway analysis confirmed NHK α_1-antitrypsin misfolding was associated with generally increased levels of ERAD and UPR elements compared with normal folding of M α_1-antitrypsin **Figure 4.6A**. In addition this approach reported a complex pattern of changes in ribosomal proteins **Figure 4.6B**, **Appendix H**) in the ER-associated proteome. Further pathway analyses of these processes are shown in **Appendix G**. However, significant alterations were again observed in proteins involved in metabolic processes and bioenergetics, **Figure 4.7**, and included contrasting changes of expression in mitochondrial and cytoplasmic pathways. Major increases were predominantly seen in glycolysis pathway components, while reductions were dominant in oxidative phosphorylation pathway components. More minor reductions were observed in proteins involved in the tricarboxylic acid cycle.

Fewer proteins are specifically downregulated in relative terms as a result of switching from expression of M to NHK α_1-antitrypsin expression and these cluster less distinctly, **Figure 4.8**. Interestingly, Rab proteins 1, 4, 5 and 14 are reduced, consistent with a reduction in ER-

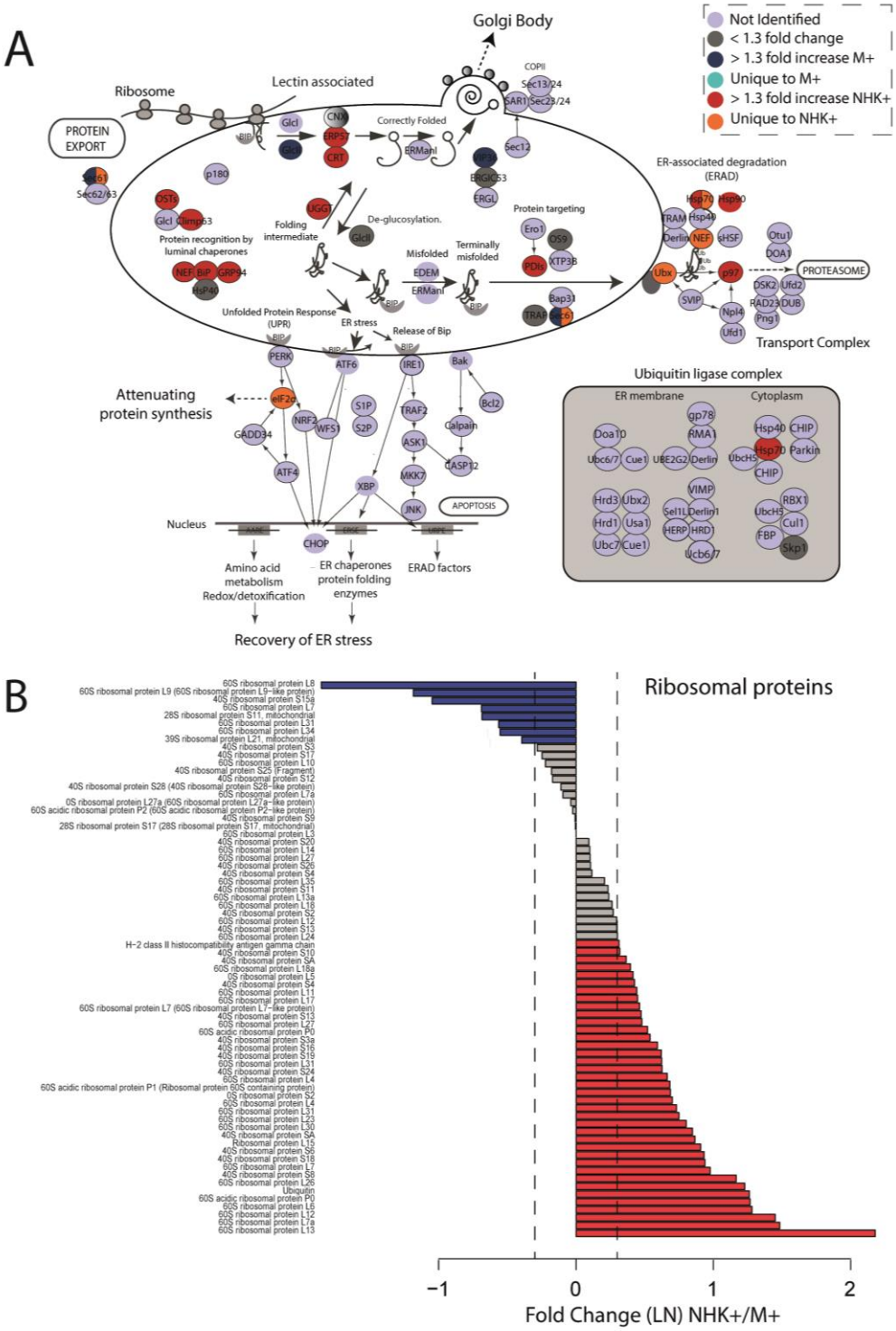


Figure 4.6 Pathway analysis (KEGG) of NHK+ expression levels to M+: ER and ribosomal processes. **A)** Functional organisation of differentially-expressed proteins annotated as belonging to ER protein processing pathways. Colours of nodes indicate measured fold change. Where more than one protein accession was mapped to the same node, more than one colour is included within the plot. **B)** Bar plot of fold changes in expression of proteins annotated as belonging to the ribosomal complex

Golgi trafficking as a result of ER stress responses. A number of subunits of mitochondrial respiratory chain complexes I, II, III and V are also downregulated in cells expressing NHK α_1-antitrypsin further pointing to alterations in bioenergetics processes, Mitochondria are key hubs of cellular metabolism and bioenergetic processes and play host to the tricarboxylic acid cycle and oxidative phosphorylation. Moreover mitochondrial dynamics are linked with alterations in redox homeostasis. Changes in bioenergetics are often associated with changes in mitochondrial function and morphology [8]. We therefore undertook live cell confocal microscopy of mitochondria *in situ* within the

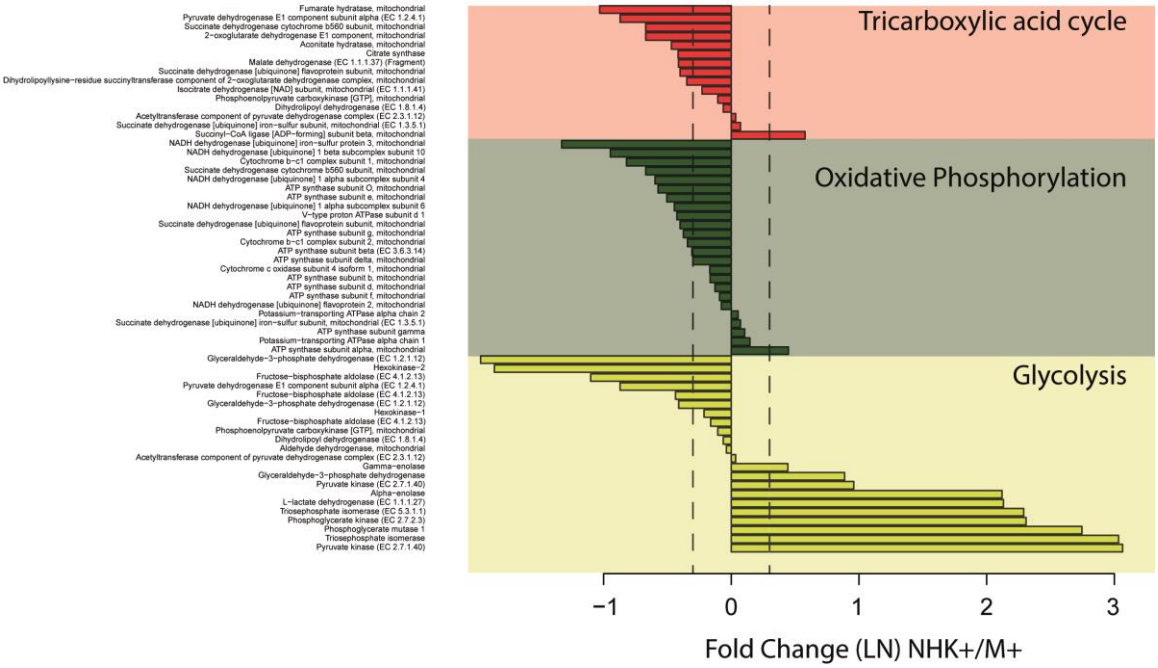


Figure 4.7 Pathway analysis (KEGG) of NHK+ expression levels to M+: Bioenergetics processes. Measured fold-change bar plot for proteins mapped to three bioenergetic pathways: the tricarboxylic acid cycle, oxidative phosphorylation and glycolysis pathways. The different pathways are affected in a non-uniform manner.

CHO-K1 cells, **Figure 4.9**. These demonstrated dramatic changes in mitochondrial morphology between cells secreting well-folded, highly-secreted M α_1-antitrypsin and those with high misfolding load of the NHK variant. Mitochondria expressing wild-type M α_1-antitrypsin were predominantly small and punctate, whilst fusiform, tubular mitochondria were prominent in the cells handling the misfolding variant. We next looked at cells expressing the more common Z variant of α_1-antitrypsin. Z α_1-antitrypsin also misfolds, but to a less dramatic degree than NHK and additionally forms polymers. The mitochondrial network in cells expressing the Z α_1-antitrypsin variant adopted a mixed

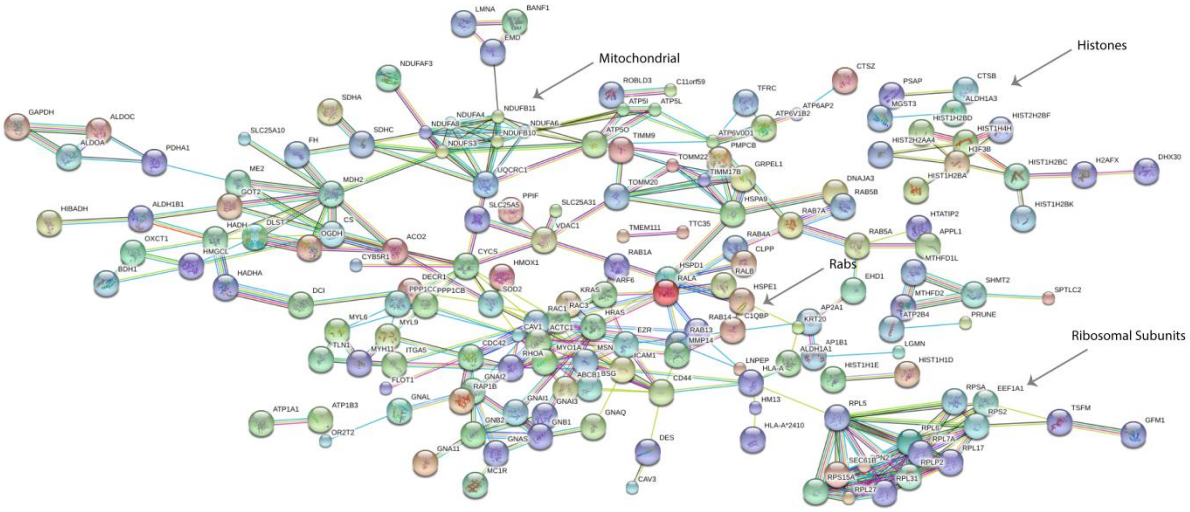


Figure 4.8 Protein-protein interaction network analysis (STRING) of proteins identified to be > 1.3 fold downregulated in cells expressing NHK α_1-antitrypsin compared with cells expressing M α_1-antitrypsin. Clusters have been manually curated to highlight networks that contain multiple proteins with similar functional roles. The interaction network is composed of interactions with at least a high confidence (Score ≥ 0.7, as defined by STRING)

morphology including both tubular and punctate structures. We were able to quantify the change in mitochondrial morphology in terms of the mitochondrial 'sphericity' in all mitochondria that could be identified in 32 cells from each cell line using unbiased, automated image processing software. In agreement with our observations, whilst variability was observed, M+, Z+ and NHK+ cells showed highly significant differences between the different cell lines ($p<0.0001$) assessed pairwise (by t-test) or by ANOVA.

Proteins that were successfully mapped during pathway analysis of the pairwise comparison of M induced cells and NHK induced cells were taken forward to include Z induced cell expression data. Proteins successfully mapped onto KEGG pathways and identified in all three conditions were analysed by unsupervised hierarchical clustering to identify expression patterns, **Figure 4.11**. Expression levels for proteins that were mapped to the ribosome complex or were identified as mitochondrion residing were found to be similar for the NHK and Z induced cells, with the two cell lines clustering together with M+ cell levels consistently being the most divergent.

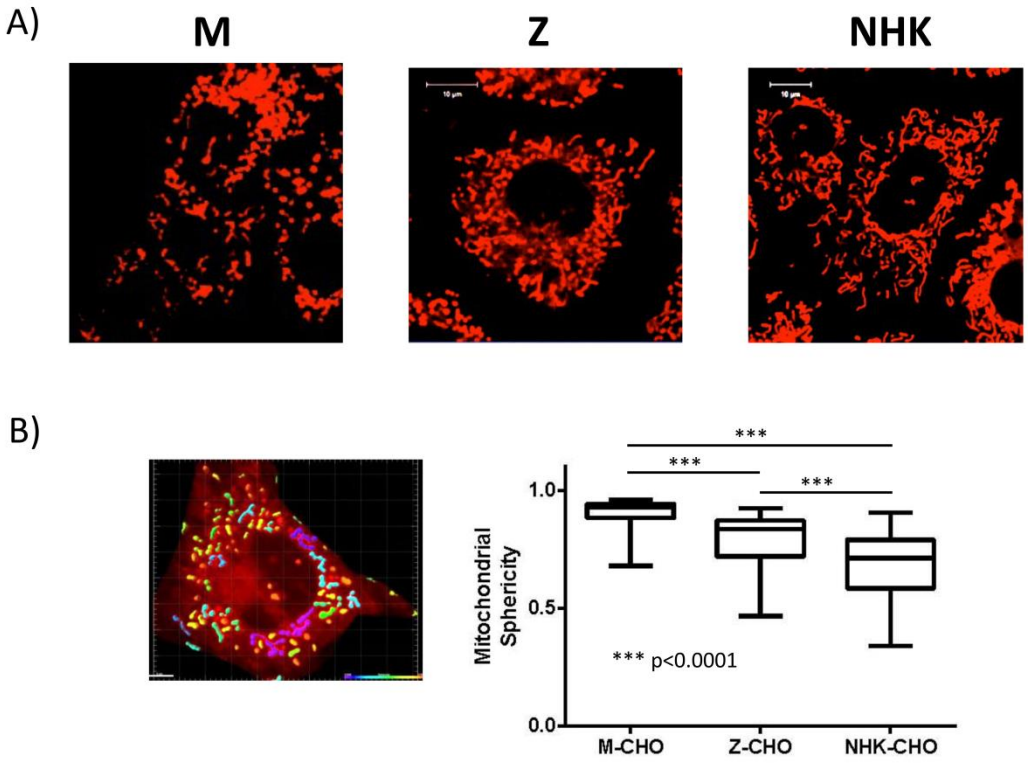


Figure 4.9 Mitochondrial morphologies associated with switching expression from well-folding to increasingly misfolded variants of α_1-antitrypsin. A: characteristic mitochondrial appearances associated with expression of each variant. B: left, example of automated analysis of mitochondrial morphology (maximal sphericity coloured entirely red and scored 1.0, more elongated structures heatmap coloured with increasing blueness and scoring lower for sphericity); right, box and whiskers (mean, interquartile range and standard errors) plot of sphericity scores for each variant. *** p<0.0001

Binary comparison did identify a subset of proteins that displayed differential expression between NHK and Z variant induction and could therefore represent signatures of unique ER . The tubulin protein, Tubulin alpha-3-chain, was identified as being the most statistically significant expression change in addition to being largest fold increase in the Z variant induced cells (> 20 upregulated). Analysis of its expression levels in the M variant induced cells found 9 times higher than NHK induced cells (P=0.0015) but not a significant difference between Z and M cells (x 3.4 up in Z, p=0.14). The detected changes are likely to be associated with the detected changes in mitochondrial morphology.

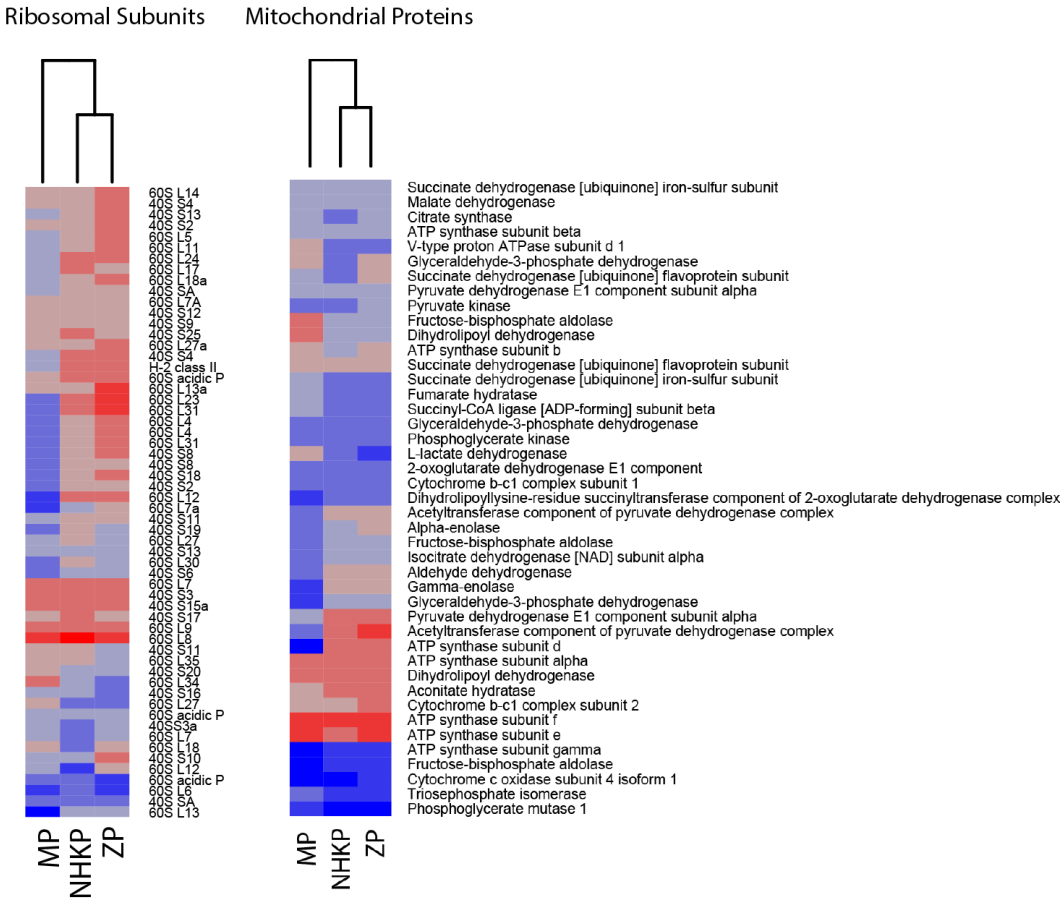


Figure 4.10 Protein intensity heatmap for expression levels detected in cell lines expressing M , NHK and Z α_1-antitrypsin. Proteins have been mapped as ribosomal proteins or mitochondrial proteins by pathway analysis. Hierarchical clustering was performed based on detected protein levels.

Proteins that had been identified as having significant difference between NHK induced and Z induced cells were analysed by gene ontology to identify biological processes that were well represented **Figure 4.11A**. Six out of seven proteins involved in translation were upregulated in Z induced cells, **Figure 4.11B**, including the elongation factor1-gamma which is responsible for the delivery of aminoacyl tRNAs to the ribosome and may be involved in tethering the assembly of multisubunit complexes. Similarly upregulated were proteins involved in oxidation reduction **Figure 4.11C**, suggesting an increase in oxidative stress upon Z variant misfolding.

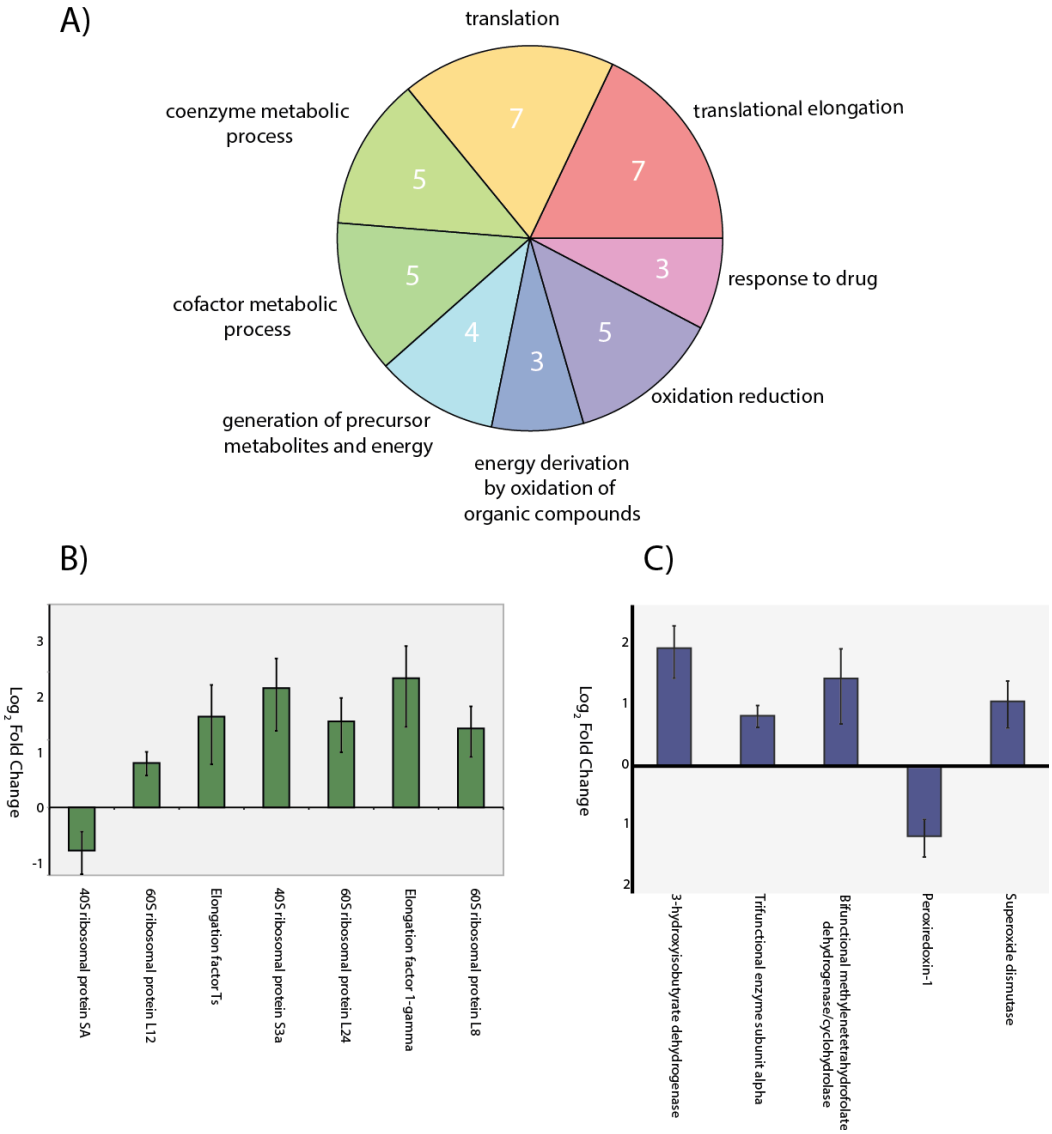


Figure 4.11 Gene ontology annotation of proteins identified as differentially expressed in pairwise comparison of NHK α1-antitrypsin and Z α1-antitrypsin induced cells. **A)** Biological process annotation of differentially expressed proteins, number of proteins mapped to each process shown in white. **B)** Measured fold change of proteins involved in translation. **C)** Measured fold change of proteins related to oxidation reduction.

4.4 Discussion

ER stress responses maximally activate protein folding quality control pathways that are of fundamental importance to eukaryote biology[33], and are increasingly implicated as common pathways activated within diverse disease states. The CHO cell system is one of the most commonly used experimental models of mammalian cells. Moreover, the synthesis and appropriate folding of high volumes of secreted proteins in CHO cells is also important to biotechnology industries [34]. M α_1-antitrypsin is one of the most highly secreted proteins in the human body, and the NHK variant is widely used as an experimental trigger of ER stress responses. While whole cell studies, including proteomic approaches can provide valuable data, ER-focused studies are enriched to assess direct responses to protein misfolding within that compartment. Our approach is further complementary to classical cell biology studies as it informs upon changes that persist after accounting for expansion of the ER compartment as a whole. While it enhances changes in terms of ER-localised proteins this approach appears similarly capable of highlighting alterations within the mitochondrial compartment, and reporting major changes in cytosolic proteins.

The changes in ER proteins that we have identified between states associated with folding of these variants are plausible based upon existing knowledge, while providing new insights. Thus it is consistent that numerous elements of the ER quality control machinery for handling misfolded polypeptides are found in higher abundance in the cells expressing NHK than M variants. Within the ER these include ERp57, calreticulin, UDP-glucose:glycoprotein glucosyltransferase (UGGT) and cytoskeleton-linking membrane protein of 63kDa (CLIMP-63) that regulates Dicer [35]. They also include the interacting chaperones BiP (Heat Shock Protein (HSP)70 family member), nuclear exchange factor (NEF), and glucose regulated protein (GRP)94 (the ER-resident HSP90 family member) that interact together. Levels of the ER-associated HSP40 protein DnaJ-like subfamily B member (DnaJB)11 that co-chaperones with BiP and GRP94 is also detected and is stable between cells expressing both M and NHK α_1-antitrypsin. As detailed above, BiP biology is critical in both chaperoning protein folding in the ER and mediating the UPR in the face of misfolding and proteostatic challenge. The coupling of BiP to the nuclear exchange factor SEL1 facilitates substrate release. However the proteins that are increased within the ER-associated proteome in the context of NHK α_1-antitrypsin expression that are mapped as NEFs (accession numbers G3I973 and G3GWF4)

appear closest to members of the HSP105/110 class of chaperones. Whilst they contain nucleotide binding domains and associate with HSP70 and HSP40 chaperones, it is not clear whether they are mainly functioning as nucleotide exchange factors. Indeed they have been recently shown to represent the mammalian functional counterpart, [36], of the HSP104/ClpB disaggregase family in yeast and bacteria. When associated with HSP70 and HSP40 family members such proteins have the ability to disassemble protein aggregates including amyloid [37], [38]. Reference to pathway analysis identified hypoxia upregulated protein 1 (3.4 fold high in NHK+, p value < 0.05) and HSP105 fragment (uniquely identified in NHK+ system) as being directly mapped to NEF within the 'Protein processing in the ER' KEGG pathway. Sequence alignment of hypoxia upregulated protein 1 against human HSP105 validated the mapping results with a sequence identity of 27%, **Figure 4.12**. Taken together, these changes in the ER-associated proteome therefore suggest that expression of NHK α_1-antitrypsin may predispose to the formation of non-polymeric aggregates that are then degraded by the BiP(HSP70):DnaJB11(HSP40):NEF(HSP110) system. Interestingly a similar response was observed upon Z α_1-antitrypsin misfolding (2.8 fold increases, p value < 0.005). Z variant is known to form well order polymer aggregates and the regulation of disaggregase associated proteins in response is not unexpected.

This pathway classification also includes proteins that act at the cytoplasmic side of the ER membrane and beyond this into the cytosol. The translation initiation factor eIF2α and some cytoplasmic ERAD response factors are observed, suggesting that proteins that interact transiently with ER membrane bound proteins may be detected by this approach. The β-subunit of cytosolic HSP90 is particularly highly upregulated in the latter category. While the detailed significance of this requires elucidation, cytosolic HSP90 is implicated in stabilisation of the UPR factor IRE1 [39].

A complex ribosomal protein response to α_1-antitrypsin misfolding is reported. This likely reflects changes in the composition of rough ER-associated ribosomes. In general translation of polypeptides into the ER is reduced in ER stress as a compensatory mechanism. The increase in many ribosomal proteins in the ER-associated proteome of cells expressing NHK α_1-antitrypsin relative to the wild-type protein was therefore unexpected. The observed changes might conceivably contribute to reduce translation rates of mRNAs encoding secreted proteins. Such selective upregulation of ER stress-modifying proteins is evident in our data reporting increases in chaperone elements. Additionally, the synthesis of some UPR targets is preserved in ER stress or, as in the case of ATF4, enhanced. The

Figure 4.12 Sequence alignment of hypoxia upregulated protein 1, as identified from the Chinese hamster protein database (Uniprot) with human heat shock protein 105.

role of altered ribosomal protein abundance has yet to be examined in these responses, but our data indicate it deserves further analysis. Another consideration is that the co-elution of mitochondrial-residing proteins, which could potentially include ribosomal proteins, may be producing the complex response observed. Mitochondrial ribosomal elements may differ from those that make up the rough ER. In order to confirm this whole cell analysis would be required. In addition to proteins that are

clearly defined as 'ER resident' these studies also indicate changes in other proteins that are associated with the ER-containing fractions but seem unlikely themselves to be physically included within or attached to the ER. These include proteins involved with mRNA processing. The observed mitochondrial changes could relate to changes in the direct interactions between mitochondria and the ER as well as changes within mitochondria. Intimate physical and functional links exist between the ER and mitochondria, including machinery related to mitochondrial dynamics such as mitofusin (Mfn)2, dynamin-related protein (Drp)1 and mammalian mitochondrial Rho (Miro)1.[40] Acute ER stress has been demonstrated to stimulate an adaptive increase in mitochondrial respiration, promoting cell survival [41]. Moreover, prolonged activation of the ER stress pathway may induce mitochondrial dysfunction and apoptosis [42]. The morphological changes we observed in mitochondria indicate that the balance of fusion and fission events in the mitochondrial network is also affected by the expression of the NHK-variant relative to wild-type M α_1-antitrypsin. We therefore hypothesise three different and potentially co-existing mechanisms may underlie this element of cross-talk between ER protein folding pathways and the dynamics of mitochondrial turnover, **Figure 4.14**. Firstly, the ER plays a physical role in at least some mitochondrial fission events [43]. The ER stress response is known to streamline its functions to relieve the proteostatic challenge, reducing polypeptide synthesis in general and increasing chaperone levels [33]. Changes in the ER proteome may therefore also reflect an altered contribution to mitochondrial fission.

Secondly the difference in mitochondrial morphology between high secretion and misfolding contexts may reflect different bioenergetic requirements within the cell. Indeed, overall the data support a shift of cellular metabolism from oxidative phosphorylation towards increased glycolysis. In addition the UPR/ER stress response alters energetic demands within the ER to focus energy utilisation upon alleviation of proteostatic load. Alterations in redox state, that may relate to the changes in levels of enzymes involved in disulphide-bond formation and glutathione processing, can also modify mitochondrial morphology [44].

Additionally, mitochondrial stress responses may play a role. Unresolved ER stress may lead to apoptosis, and typically involves autophagy, including mitochondrial autophagy (mitophagy). During mitophagy, damaged mitochondria are incorporated into autophagosomes and targeted for degradation. However under conditions of cellular stress, healthy mitochondria may undergo an adaptive fusion response, sparing them from autophagic degradation and promoting cell survival.

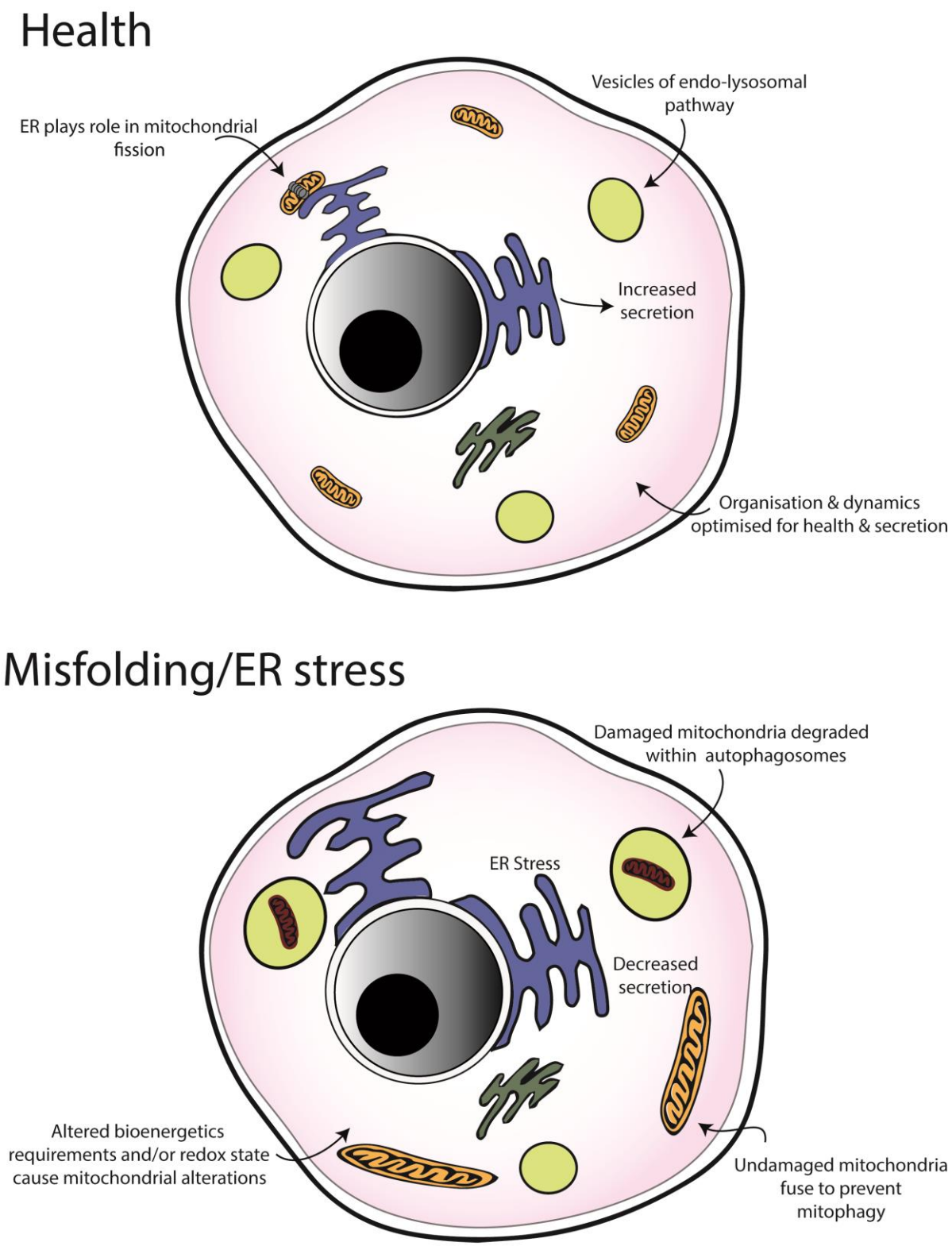


Figure 4.13 Potential mechanisms for alterations in mitochondrial morphology with α_1-antitrypsin misfolding vs secretion

Such alterations in mitochondrial morphology are consistent with our observations of a shift towards elongated mitochondria accompanying chronic ER-stress in NHK-variant expressing cells. It is possible that the changes in mitochondrial morphology also relate to the associated increases in cytoskeletal elements that we observe.

To explore the wider disease relevance of this process we also imaged mitochondria in the CHO cell model of Z (Glu342Lys) α_1-antitrypsin expression. This is a far more common deficiency variant than NHK α_1-antitrypsin, its clinicopathological consequences have been studied more extensively. The Z variant misfolds to a lesser degree than NHK α_1-antitrypsin with ~70% of synthesised Z α_1-antitrypsin undergoing degradation by ERAD in cell models [45]. However around 15% of the synthesised Z α_1-antitrypsin polypeptides fold sufficiently to self-associate into ordered chains of α_1-antitrypsin polymers. In hepatocytes, and in CHO cells these are retained within the ER. The remaining material is secreted. Polymer accumulation within inclusion bodies is believed to mediate hepatic toxicity. It is associated with a lowered threshold to ER stress in the context of a second stimulus, but does not directly activate the UPR in CHO cells or hepatocytes. The CHO cell model recapitulates the ER accumulation of Z α_1-antitrypsin (~15%) that avoids ERAD, but has aberrant conformational behaviour resulting in self-association to form polymer chains rather than secretion. The mitochondrial phenotype has not been characterised in these cells to date. However, *ex vivo* samples of hepatic tissue from PiZ mice and human patients demonstrate evidence of an increased mitochondrial autophagy response.[46] Moreover, studies using a hepatoma cell line expressing Z α_1-antitrypsin demonstrated a mitochondrial damage response that could be abrogated by boosting ERAD. These data suggested that proteostatic load and/or polymerisation contributed to mitochondrial damage and mitophagy in Z α_1-antitrypsin deficiency. ER accumulation of serpin polymers is known to trigger specific ER responses rather than the classical UPR, suggesting that the presence of polymers directs subcellular responses independently of protein misfolding *per se*. Nevertheless our data demonstrate a correlation of genotype to mitochondrial phenotype from PiM to PiZ to PiNull$_{HongKong}$ that matches the degree of proteostatic challenge rather than the potential to form polymers. Specifically mitochondria in cells expressing Z α_1-antitrypsin showed more mixed fusiform and punctate morphologies, giving sphericity scores intermediate between the M and NHK expression models, but significantly different from either.

These phenotypic changes may have potential as live cell readouts of α_1-antitrypsin misfolding versus folding and secretion, amenable to automated quantification in high throughput compound screens for drug discovery. This would first require specific validation for this application. In particular, studies would need to ascertain whether shifts of mitrochondrial morphology in response to treatment did in fact correlate well with reduced proteostatic challenge and/or dysfunctional consequences. In cell models of Z α_1-antitrypsin deficiency, it would also be important to assess how such changes correlated with alterations in ER polymer load since these appear to have toxic properties relevant to the development of liver disease.

Reductionist studies have been used to good effect to characterise the cellular responses to NHK and Z α1-antitrypsin misfolding including those unique to each variant. Proteomic analysis allows the identification of cellular responses in a more unbiased and high-throughput manner. In this study we hoped to use the approach to better define the differences in ER specific responses. Surprisingly our data only identified a small subset of proteins that were significantly changed between the two variants. Stress responses induced upon NHK misfolding and detected by our workflow were on the whole recapitulated in cells expressing the Z variant. Our data therefore suggests that NHK and Z α1-antitrypsin misfolding promotes ER stress responses to a more similar extent than reductionist studies have found. Mass spectrometry-based proteomic studies, despite continued improvements in instrument speed and sensitivity are still limited in the depth proteins that can be quantified from a biological sample and this study is no exception. As the ERAD response is the primary response element to unfolding in the ER and is known to be common between NHK and Z variants, expression changes induced by the ERAD may be preferentially detected. Being the primary unfolding response, it would be expected that many of the proteins directly involved in the ERAD response would be highly expressed. Indeed based upon Hi3 intensity BIP is the second most highly expressed protein detected in the NHK variant induced cells. Our results therefore may be a result of inadequate dynamic range to probe other ER specific cellular responses that may involve lower abundance proteins. The use of new instrumentation with higher sensitivity and dynamic range, increased protein level fractionation or a combination of both would help improve the depth of coverage and help better identify cellular responses unique to the misfolding of the two variants.

Although limited, a small set of proteins were identified as differentially regulated between the two conditions. Tubulin alpha-3-chain, the most significant of the changes detected could be related to

mitochondrial morphology changes however expression levels better correlate to the expressed variants ability to form ER residing polymers (8-fold higher NHK+, 26-fold higher Z+). This raises the possibility that Tubulin alpha-3-chain, which is known to be involved in protein polymerization directly associates with the polymer form of α_1-antitrypsin. Another notable upregulated protein was calumenin, an ER residiing calcium-binding protein. It is known to play a role in protein folding and protein sorting and importantly has been shown in mice to alleviate ER stress as monitored by reduced levels of ER stress markers GRP78, PERK and CHOP. Calumenin is upregulated in response to NHK α1-antitrypsin misfolding(3-fold increase) but is upregulated further still in response to Z variant expression (4-fold increase). Our data therefore suggests that calumenin could be a marker for the ER overload response which is known to be activated to a far higher degree by Z variant misfolding.

Proteins involved in translation, including both ribosomal proteins and translation initiation factors were in general found to be more abundant in cells expressing Z α1-antitrypsin. Interestingly one these proteins EIF1γ is involved in protein assembly and folding as it possess a glutathione-S-transferase domain. Why the formation of Z polymer would induced this change is unclear.

Oxidative reduction-related proteins showed a similar trend for upregulation in Z variant inducing cells. Data available for other misfolding diseases, such as Alzheimer's disease, provides evidence towards one of the toxic mechanisms of protein aggregation being the generation of reactive oxygen species (ROS). Upregulation of proteins involved in oxidative reduction could represent a protective mechanism induced upon Z polymer formation. The upregulation of superoxide dismutase in particular could be a major protective element. It has been shown previously in a mouse Alzheimer's disease model that the overexpression of SOD-2, reduces AD pathology.

4.5 Conclusion

An ER focused label-free quantitative proteomic workflow has been developed to obtain expression data on ER stress responses to the misfolding of two α_1-antitrypsin variants relative to basal levels upon wild-type α_1-antitrypsin expression in CHO cells. Overall, our findings validate a focused, systematic approach that provides a general context for understanding proteomic correlates of misfolding responses as defined by classical reductionist cell biological studies in this widely used model of ER stress. Expected responses, including the upregulation of the well characterised molecular chaperones upon NHK α1-antitrypsin induction were quantified. Further to this, previously uncharacterised ER responses were quantified, including a complex ribosomal expression pattern that represents a new resource for identifying ER stress related targets in reductionist studies. Interestingly NEF, a member of the HSP 105/110 family which has been implicated as a functional counterpart of the bacterial HSP104 disaggregase was found upregulated in both NHK and Z α_1-antitrypsin induced cells. 15% of expressed Z α1-antitrypsin is known to form ordered polymers in hepatocyte ER, however polymers are not observed for NHK α_1-antitrypsin. Our findings present the possibility that NEF as part of the HSP70-HSP40-HSP110 system are able to dissemble early non-polymeric aggregates. Such findings demonstrate that this approach can provide novel mechanistic insights relevant to misfolding diseases.

A known limitation of the use of a gradient density for ER residing proteins is the co-elution of organelles with similar densities, namely the mitochondria. Despite being a known contaminant, the large observed fold change measurements in the data lead us to conduct live cell confocal imaging of mitochondrial was conducted on CHO cells expressing the three α_1-antitrypsinvariants. Mitochondrial morphological changes, which have not been previous reported in the context of ER stress, were observed when the two misfolding variants were expressed. Interestingly a statistically significant difference was observed between all three cell lines, indicating the degree/type of misfolding effects mitochondrial response. We propose three possible mechanisms that could change the balance of mitochondrial fission/fusion. Firstly streamlining of ER functions upon increasing proteostasis challenge may reduce the role it plays in mitochondrial fission events. Secondly changes in the balance of secretion and misfolding may present different bioenergetic cellular requirements, which are subsequently represented by changes in mitochondrial fission/fusion events. Lastly, chronic ER stress is known to trigger apoptosis including autophagy. It has been previously demonstrated that

mitochondrial elongation is able to prevent mitophagy and promote cell survival [47] Mitochondrial morphological changes could be a result of an adaptive fusion response to increased apoptosis. Interestingly we observed the degree of mitochondrial elongation was directly relatable to the degree of protein misfolding but does not appear to be affected by protein polymerisation. These findings are amenable to further interrogation and hypothesis refinement in future studies. A major element of such work will be to define the proteins involved in processes of molecular cross-talk between protein misfolding within the ER and mitochondrial behavior. These may relate to the disease mechanism and/or ameliorative responses in α_1-antitrypsin deficiency. If indeed these morphological changes are validated, most likely in a further range of α_1-antitrypsin variants, an interesting prospect would the use of our findings as a relatively fast functional read out for drug binding studies.

Recently it was demonstrated that an approach integrating ion mobility mass spectrometry and nuclear magnetic resonance (NMR) can be used to screen small-molecule ligands that confer thermodynamic stability to α_1-antitrypsin misfolding variants [48]. By monitoring conformational flexibility upon ligand binding in the presence of increasing collisional energy the efficacy of small molecules can potentially be screened in a high throughput manner. By the further incorporation of a measure of mitochondrial morphology into the screen approach, the in vivo effects of ligand binding on the extent of protein misfolding could be reported. It would be expected that the introduction of this extra information should increase the confidence of small molecule drug efficacy.

Limited differences in ER responses were detected between the two misfolding variants, Z and NHK α_1-antitrypsin. This is most likely to be a combination of the ERAD response being the predominant protein signature being activated by the induction of misfolding α_1-antitrypsin variants and the limited coverage achieved by our instrumentation. Future studies with current generation technology and, if necessary extensive fractionation, will increase proteome coverage to an extent that close to complete coverage of the ER proteome can be characterised. Using such an approach lower abundance stress response such as the ER overload response could be successfully characterised to an extent not previously achieved. More extensive coverage would provide better information for correlating observed mitochondrial morphological alterations with ER stress signatures and should aid our understanding of its biological significance.

Proteins involved in translation and oxidative reduction were well represented in the subset of proteins that were differentially regulated. Increases in proteins involved in oxidative reduction are seen in

other misfolding diseases where polymer aggregates are formed and could suggest the aggregation of Z α_1-antitrypsin is associated with an increase in reactive oxidative species. Oxidative species are known to cause cellular damage in a wide variety of disease contexts and increased levels can even result in cell death. Our findings may present a pathological pathway from α_1-antitrypsin polymer formation to in vivo hepatocyte damage in A1ATD.

4.6 References

[1] M. Nyon and B. Gooptu, "Therapeutic targeting of misfolding and conformational change in α1-antitrypsin deficiency," *Future Med. Chem.*, vol. 6, pp. 1047–1065, 2013.

[2] J. A. Huntington, R. J. Read, and R. W. Carrell, "Structure of a serpin-protease complex shows inhibition by deformation.," *Nature*, vol. 407, no. 6806, pp. 923–926, 2000.

[3] E. Kelly, C. M. Greene, T. P. Carroll, N. G. McElvaney, and S. J. O'Neill, "Alpha-1 antitrypsin deficiency," *Respir. Med.*, vol. 104, no. 6, pp. 763–772, 2010.

[4] R. N. Sifers, S. Brashears-Macatee, V. J. Kidd, H. Muensch, and S. L. Woo, "A frameshift Mutation Results in a Truncated a1-Antitrypsin That is Retained within the Rough Endoplasmic Reticulum," *J.Biol.Chem.*, vol. 263, no. 15, pp. 7330–7335, 1988.

[5] Y. Kimata, Y. I. Kimata, Y. Shimizu, H. Abe, I. C. Farcasanu, M. Takeuchi, M. D. Rose, and K. Kohno, "Genetic Evidence for a Role of BiP / Kar2 That Regulates Ire1 in Response to Accumulation of Unfolded Proteins," *Mol. Biol. Cell*, vol. 14, pp. 2559–2569, 2003.

[6] S. Pan, S. Wang, B. Utama, L. Huang, N. Blok, M. K. Estes, K. W. Moremen, and R. N. Sifers, "Golgi localization of ERManI defines spatial separation of the mammalian glycoprotein quality control system.," *Mol. Biol. Cell*, vol. 22, no. 16, pp. 2810–2822, 2011.

[7] Y. Wu, M. T. Swulius, K. W. Moremen, and R. N. Sifers, "Elucidation of the molecular logic by which misfolded alpha 1-antitrypsin is preferentially selected for degradation.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 14, pp. 8229–8234, 2003.

[8] Y. Zhong, H. Shen, Y. Wang, Y. Yang, P. Yang, and S. Fang, "Identification of ERAD components essential for dislocation of the null Hong Kong variant of α-1-antitrypsin (NHK)," *Biochem. Biophys. Res. Commun.*, vol. 458, no. 2, pp. 424–428, 2015.

[9] E. J. Greenblatt, J. A. Olzmann, and R. R. Kopito, "Derlin-1 is a rhomboid pseudoprotease required for the dislocation of mutant α-1 antitrypsin from the endoplasmic reticulum," *Nat. Struct. Mol. Biol.*, vol. 18, no. 10, pp. 1147–1152, 2011.

[10] Y. Iida, T. Fujimori, K. Okawa, K. Nagata, I. Wada, and N. Hosokawa, "SEL1L protein critically determines the stability of the HRD1-SEL1L Endoplasmic Reticulum-associated Degradation (ERAD) complex to optimize the degradation kinetics of ERAD substrates," *J. Biol. Chem.*, vol. 286, no. 19, pp. 16929–16939, 2011.

[11] A. Schäfer and D. H. Wolf, "Sec61p is part of the endoplasmic reticulum-associated degradation machinery.," *EMBO J.*, vol. 28, no. 19, pp. 2874–2884, 2009.

[12] S. Pan, X. Cheng, and R. N. Sifers, "Golgi-situated endoplasmic reticulum α-1, 2-mannosidase contributes to the retrieval of ERAD substrates through a direct interaction with γ-COP.," *Mol. Biol. Cell*, vol. 24, no. 8, pp. 1111–21, 2013.

[13] M. J. Iannotti, L. Figard, A. M. Sokac, and R. N. Sifers, "A golgi-localized mannosidase (man1b1) plays a nonenzymatic gatekeeper role in protein biosynthetic quality control," *J. Biol. Chem.*, vol. 289, no. 17, pp. 11844–11858, 2014.

[14] S. J. Marciniak and D. Ron, "Endoplasmic Reticulum Stress Signaling in Disease," *Physiol Rev*, vol. 86, no. 4, pp. 1133–1149, 2006.

[15] M. J. Davies and D. A. Lomas, "The molecular aetiology of the serpinopathies," *Int. J. Biochem. Cell Biol.*, vol. 40, no. 6–7, pp. 1273–1286, 2008.

[16] K. D. Fairbanks and A. S. Tavill, “Liver disease in alpha 1-antitrypsin deficiency: a review.,” *Am. J. Gastroenterol.*, vol. 103, no. 8, pp. 2136–2141, 2008.

[17] A. M. Wood and R. A. Stockley, “Alpha one antitrypsin deficiency: From gene to treatment,” *Respiration*, vol. 74, no. 5, pp. 481–492, 2007.

[18] M. W. Lawless, C. M. Greene, A. Mulgrew, C. C. Taggart, S. J. O’Neill, and N. G. McElvaney, “Activation of endoplasmic reticulum-specific stress responses associated with the conformational disease Z alpha 1-antitrypsin deficiency.,” *J. Immunol.*, vol. 172, no. 9, pp. 5722–5726, 2004.

[19] H. L. Pahl and P. A. Baeuerle, “The ER-overload response: activation of NF-kappa B.,” *Trends Biochem. Sci.*, vol. 22, no. 2, pp. 63–67, 1997.

[20] T. P. J. Dunkley, R. Watson, J. L. Griffin, P. Dupree, and K. S. Lilley, “Localization of organelle proteins by isotope tagging (LOPIT),” *Mol. Cell. Proteomics*, vol. 3, no. 11, pp. 1128–1134, 2004.

[21] M. P. Rout and M. C. Field, “Isolation and characterization of subnuclear compartments from Trypanosoma brucei. Identification of a major repetitive nuclear lamina component,” *J. Biol. Chem.*, vol. 276, no. 41, pp. 38261–38271, 2001.

[22] C. Pasquali, I. Fialka, and L. A. Huber, “Subcellular fractionation, electromigration analysis and mapping of organelles,” *J. Chromatogr. B Biomed. Sci. Appl.*, vol. 722, no. 1–2, pp. 89–102, 1999.

[23] S. E. Stimpson, J. R. Coorssen, and S. J. Myers, “Optimal isolation of mitochondria for proteomic analyses,” *Anal. Biochem.*, vol. 475, pp. 1–3, 2015.

[24] A. Ordóñez, E. L. Snapp, L. Tan, E. Miranda, S. J. Marciniak, and D. A. Lomas, “Endoplasmic reticulum polymers impair luminal protein mobility and sensitize to cellular stress in alpha1-antitrypsin deficiency,” *Hepatology*, vol. 57, no. 5, pp. 2049–2060, 2013.

[25] M. Wilm, A. Shevchenko, T. Houthaeve, S. Breit, L. Schweigerer, T. Fotsis, and M. Mann, “Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry.,” *Nature*, vol. 379, no. 6564, pp. 466–469, 1996.

[26] S. J. Geromanos, J. P. C. Vissers, J. C. Silva, C. A. Dorschel, G.-Z. Li, M. V Gorenstein, R. H. Bateman, and J. I. Langridge, “The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS.,” *Proteomics*, vol. 9, no. 6, pp. 1683–95, 2009.

[27] J. C. Silva, M. V Gorenstein, G.-Z. Li, J. P. C. Vissers, and S. J. Geromanos, “Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition.,” *Mol. Cell. Proteomics*, vol. 5, pp. 144–156, 2006.

[28] V. J. Patel, K. Thalassinos, S. E. Slade, J. B. Connolly, A. Crombie, J. C. Murrell, and J. H. Scrivens, “A comparison of labeling and label-free mass spectrometry-based proteomics approaches.,” *J. Proteome Res.*, vol. 8, no. 7, pp. 3752–9, Jul. 2009.

[29] D. W. Huang, B. T. Sherman, and R. A. Lempicki, “Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.,” *Nat. Protoc.*, vol. 4, no. 1, pp. 44–57, 2009.

[30] A. Franceschini, D. Szklarczyk, S. Frankild, M. Kuhn, M. Simonovic, A. Roth, J. Lin, P. Minguez, P. Bork, C. von Mering, and L. J. Jensen, “STRING v9.1: protein-protein interaction networks, with increased coverage and integration.,” *Nucleic Acids Res.*, vol. 41, pp. D808–15, Jan. 2013.

[31] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, "Cytoscape : A Software Environment for Integrated Models of Biomolecular Interaction Networks Cytoscape : A Software Environment for Integrated Models of Biomolecular Interaction Networks," *Genome Res.*, vol. 12, no. 11, pp. 2498–2504, 2003.

[32] J. Cai, G. Chen, and J. Wang, "ClusterViz: a Cytoscape plugin for graph clustering and visualization," *Sch. Inf. Sci. Eng.*, pp. 2–4, 2010.

[33] J. E. Chambers and S. J. Marciniak, "Cellular Mechanisms of Endoplasmic Reticulum Stress Signaling in Health and Disease. 2. Protein Misfolding and ER stress," *Am. J. Physiol. Cell Physiol.*, vol. 307, pp. C657–C670, 2014.

[34] Y. J. Kim, E. Baek, J. S. Lee, and G. M. Lee, "Autophagy and its implication in Chinese hamster ovary cell culture," *Biotechnol. Lett.*, vol. 35, no. 11, pp. 1753–1763, 2013.

[35] G. Pépin, M. P. Perron, and P. Provost, "Regulation of human Dicer by the resident ER membrane protein CLIMP-63," *Nucleic Acids Res.*, vol. 40, no. 22, pp. 11603–11617, 2012.

[36] M. L. Duennwald, A. Echeverria, and J. Shorter, "Small heat shock proteins potentiate amyloid dissolution by protein disaggregases from yeast and humans," *PLoS Biol.*, vol. 10, no. 6, 2012.

[37] J. R. Glover and S. Lindquist, "Hsp104, Hsp70, and Hsp40: A novel chaperone system that rescues previously aggregated proteins," *Cell*, vol. 94, no. 1, pp. 73–82, 1998.

[38] M. Carroni, E. Kummer, Y. Oguchi, P. Wendler, D. K. Clare, I. Sinning, J. Kopp, A. Mogk, B. Bukau, and H. R. Saibil, "Head-to-tail interactions of the coiled-coil domains regulate ClpB activity and cooperation with Hsp70 in protein disaggregation," *Elife*, vol. 2014, no. 3, pp. 1–22, 2014.

[39] M. G. Marcu, M. Doyle, A. Bertolotti, D. Ron, L. Hendershot, and L. Neckers, "Heat shock protein 90 modulates the unfolded protein response by stabilizing IRE1alpha," *Mol. Cell. Biol.*, vol. 22, no. 24, pp. 8506–8513, 2002.

[40] A. A. Rowland and G. K. Voeltz, "Endoplasmic reticulum–mitochondria contacts: function of the junction," *Nat. Rev. Mol. Cell Biol.*, vol. 13, no. 10, pp. 607–625, 2012.

[41] R. Bravo, J. M. Vicencio, V. Parra, R. Troncoso, J. P. Munoz, M. Bui, C. Quiroga, A. E. Rodriguez, H. E. Verdejo, J. Ferreira, M. Iglewski, M. Chiong, T. Simmen, A. Zorzano, J. A. Hill, B. A. Rothermel, G. Szabadkai, and S. Lavandero, "Increased ER-mitochondrial coupling promotes mitochondrial respiration and bioenergetics during early phases of ER stress.," *J. Cell Sci.*, vol. 124, no. Pt 13, pp. 2143–2152, 2011.

[42] G. Szabadkai and R. Rizzuto, "Participation of endoplasmic reticulum and mitochondrial calcium handling in apoptosis: More than just neighborhood?," *FEBS Lett.*, vol. 567, no. 1, pp. 111–115, 2004.

[43] J. R. Friedman, L. L. Lackner, M. West, J. R. DiBenedetto, J. Nunnari, and G. K. Voeltz, "ER Tubules Mark Sites of Mitochondrial Division," *Science*, vol. 334, no. 6054, pp. 358–362, 2011.

[44] P. H. G. M. Willems, R. Rossignol, C. E. J. Dieteren, M. P. Murphy, and W. J. H. Koopman, "Redox Homeostasis and Mitochondrial Dynamics," *Cell Metab.*, vol. 22, no. 2, pp. 207–218, 2015.

[45] H. Kroeger, E. Miranda, I. MacLeod, J. Pérez, D. C. Crowther, S. J. Marciniak, and D. A. Lomas, "Endoplasmic reticulum-associated degradation (ERAD) and autophagy cooperate to degrade polymerogenic mutant serpins," *Journal of Biological Chemistry*, vol. 284, no. 34. pp. 22793–22802, 2009.

[46] J. H. Teckman, J.-K. An, K. Blomenkamp, B. Schmidt, and D. Perlmutter, “Mitochondrial autophagy and injury in the liver in alpha 1-antitrypsin deficiency.,” *Am. J. Physiol. Gastrointest. Liver Physiol.*, vol. 286, no. 5, pp. G851–G862, 2004.

[47] L. C. Gomes, G. Di Benedetto, L. Scorrano, "During autophagy mitochondria elongate, are spared from degradation and sustain cell viability", *Nat Cell Biol,* vol. 13, no. 5, pp 589-98, 2011

[48] M. P. Nyon, T. Prentice, J. Day, J. Kirkpatrick, G. N. Sivalingam, G. Levy, I. Haq, J. A. Irving, D. A. Lomas, J. Christodoulou, B. Gooptu, K. Thalassinos, "An integrative approach combining ion mobility mass spectrometry, X-ray crystallography, and nuclear magnetic resonance spectroscopy to study the conformational dynamics of α_1-antitrypsin upon ligand binding," *Protein Science,* vol. 24, no. 8, pp 1301-1312.

5.0 Conclusions

5.1 Project Summary

Correct protein folding is essential for proper cellular function. Numerous protective mechanisms exist for maintaining protein fold homeostasis [1], [2]. Despite this the misfolding, oligomerisation and subsequent accumulation of proteins within tissue is common to a wide range of chronic diseases. Misfolding diseases account for the majority of neurodegenerative disorders and many of these are associated with ageing. The link between protein misfolding and ageing is, however, not fully understood. With an ever increasingly aged population the socioeconomic burden of increased neurodegenerative disease prevalence is poised to escalate [3], [4]. Better understanding of the association between ageing and neurodegeneration is key to developing treatments for late-onset Alzheimer's disease. Ageing, however, is not the only factor that can contribute to the development of misfolding diseases. Indeed polymorphisms have been identified that lead to a number of neurodegenerative diseases [5], [6]. The serpinopathies, which includes α_1-antitrypsin deficiency (AATD), is a class of disease caused by the misfolding of a serpin protein [7]. The serpinopathies are of interest because genetic variants that induce gross misfolding represent the aetiology of the disease.

Label-free quantification is emerging as a powerful mass spectrometry-based proteomic tool for studying complex disease states. Improvements in mass spectrometry sensitivity and resolving power, alongside improved liquid chromatography peak capacity and robustness now means that label-free quantification precision is now closely comparable to label-based methods [8], [9]. Benefits of label-free quantification include ease of implementation, low relative cost and importantly the ability to compare an unlimited number of samples.

Initially quantitative proteomics, GeLC-MS^E analysis, was undertaken as a follow-up, unbiased, approach to investigate previously published data demonstrating an age-related vulnerability to Aβ42 toxicity (Chapter 2). The experimental design allowed the identification of differentially expressed proteins in the context of ageing and Aβ toxicity. The approach successfully identified proteins previously implicated in AD but provided previously unknown information on protein specific

responses to ageing in the presence of Aβ42 toxicity. Data identified chaperoning proteins and small GTPase-related proteins as interesting biological targets for follow-up functional studies. Based on results from this project, *D. melanogaster* genetic lines have been obtained to identify the functional consequence of differentially expressed proteins, initially in terms of lifespan and climbing phenotypes.

Since its development, and subsequent integration into a commercial instrument, ion mobility mass spectrometry (IMMS), has become an effective tool for studying protein and protein complex geometries [10]–[12]. Recently the separating capacity of IM was extended to proteomic analysis and has been shown to increase protein identification rates > 50% without an increase in analysis run time. Using this new technology, $HDMS^E$, a second orthogonal project was undertaken to better understand the link between ageing and AD. The increased identification rates afforded by $HDMS^E$ in comparison to MS^E allowed a more comprehensive comparison of wild-type ageing and chronic Aβ induction (Chapter 3). Quantitative time-course analysis of *D. melanogaster* ageing with and without the induction of an aggregating variant of Aβ42 successfully identified differentially expressed proteins that were attributed to either ageing, Aβ42 toxicity or resulted from a combination of both conditions. Cluster analysis, gene ontology enrichment and protein-protein interaction analysis were all used to consider the biological significance of the quantitative data. Alterations in mitochondria, metabolic function, translation control and detoxification pathways were all well represented in our data. Pleasingly parts of our data were successfully validated by comparison with the previous proteomic approach (Chapter 1), recently published proteomic data [13] and recently published transcriptomic data [14].

Protein misfolding is known to activate a variety of cellular stress responses. AATD's activation of different endoplasmic reticulum (ER) stress responses has been previously characterised, in-part, by reductionist studies. Using the increased coverage provided by label-free quantification we used GeLC-MS^E analysis, combined with subcellular enrichment of the ER, to characterise the ER enriched response to the misfolding of two α_1-antitrypsin variants in Chinese hamster ovary (CHO) cells. Data were directly compared to the ER response of the expression of the well-folded wild-type variant (Chapter 4). Stress responses quantified included those known to participate in the ER associated degradation (ERAD) and unfolded protein response (UPR) pathways. Interestingly, co-fractionating

proteins associated with bioenergetics were found to display large fold changes. *In vivo* confocal imaging of mitochondria in the presence of misfolded α_1-antitrypsin successfully correlated the observed fold changes with mitochondrial morphological alterations. The results from this project represent novel findings and implicate mitochondrial dysregulation upon chronic ER stress.

5.2 Future Directions

Mass Spectrometry-based proteomics has, in recent years, seen dramatic improvements in acquisition speed and sensitivity. The identification and quantification of whole proteomes is no longer the impossible task it once seemed [15], [16]. This, alongside the continued developments in label-free quantification, which can be applied to any sample type, means that mass spectrometry has cemented its position as the most effective analytical technique for dissecting complex cellular processes and systems.

One particular area that has seen recent development is data independent acquisition (DIA), which aims to reduce the intensity bias that is inherent in traditional data dependent acquisition (DDA) methods [17]–[19]. Due to the high levels of precision and its ability to quantify large numbers of proteins in relatively short run times, SWATH in particular, is being heavily adopted by biomarker discovery labs as a replacement for immunoassay techniques during the validation stages of the discovery workflow. This is a good example of the increasingly large push towards replacing traditional immunoassay-based validation techniques such as Western blotting with mass spectrometry-based assays [20]. Indeed acquisition modes including SRM, pseudo SRM, PRM and now DIA offer higher sensitivity, precision and specificity than Western blotting assays. The ease of DIA implementation and the increasing number of software tools being developed to aid quantification means we could soon begin to see the routine adoption of mass spectrometry as the method of choice for protein quantification validation.

Despite improvements in mass spectrometry technologies, the challenges facing discovery proteomics are still large. Label-free quantification has helped increase the amount of information that can be obtained from analytical runs but the data is only capture a subset of the proteome present in a sample [21]. The development of label-free methods for absolute quantification is a perfect example of where mass spectrometry can now, with relative ease, provide accurate quantitative data where previously the assay would have been complex and expensive [22], [23]. A recent paper, however, demonstrated how much information is still discarded from most discovery proteomic experiments. Work by J.M. Chick and co-workers demonstrated, with the use of a mass-tolerant database search approach, that the vast amount of unmatched MS/MS spectra originate from unidentified post translational modifications (PTMs) [24]. This novel approach to data analysis was able to identify large

numbers of PTMs and amino acid variants but was still unable to assign all spectra with confident identifications. As mass spectrometry-based proteomics moves towards a meta-analysis approach of cataloguing proteomes and mass spectrometry data becomes ever more comprehensive, novel computational methods such as those by J.M Chick and colleagues will be an increasingly important tool towards the routine identification and quantification of all PTMs, single nucleotide polymorphisms (SNPs) and other chemical variants present.

5.3 References

[1] R. I. Morimoto and A. M. Cuervo, “Protein homeostasis and aging: taking care of proteins from the cradle to the grave.,” *J. Gerontol. A. Biol. Sci. Med. Sci.*, vol. 64, no. 2, pp. 167–70, 2009.

[2] H. Koga, S. Kaushik, and A. M. Cuervo, “Protein homeostasis and aging: The importance of exquisite quality control.,” *Ageing Res. Rev.*, vol. 10, no. 2, pp. 205–15, 2011.

[3] R. Brookmeyer, E. Johnson, K. Ziegler-Graham, and H. M. Arrighi, “Forecasting the global burden of Alzheimer’s disease.,” *Alzheimers. Dement.*, vol. 3, no. 3, pp. 186–91, 2007.

[4] A. Association, “2010 Alzheimer’s disease facts and figures.,” *Alzheimers. Dement.*, vol. 6, no. 2, pp. 158–94, 2010.

[5] M. H. Polymeropoulos, C. Lavedan, E. Leroy, S. E. Ide, A. Dehejia, A. Dutra, B. Pike, H. Root, J. Rubenstein, R. Boyer, E. S. Stenroos, S. Chandrasekharappa, A. Athanassiadou, W. G. Johnson, A. M. Lazzarini, R. C. Duvoisin, G. Di Iorio, L. I. Golbe, R. L. Nussbaum, E. Reports, C. Lavedant, E. Leroyt, T. Papapetropoulos, G. Di, and L. Golbe, “Mutation in the alpha-Synuclein Gene Identified in Families with Parkinson ’ s Disease,” *Science (80-.).*, vol. 276, no. June, pp. 2045–2048, 2012.

[6] P. Hollingworth, D. Harold, R. Sims, A. Gerrish, J.-C. Lambert, M. M. Carrasquillo, R. Abraham, M. L. Hamshere, J. S. Pahwa, V. Moskvina, K. Dowzell, N. Jones, A. Stretton, C. Thomas, A. Richards, D. Ivanov, C. Widdowson, J. Chapman, S. Lovestone, J. Powell, P. Proitsi, M. K. Lupton, C. Brayne, D. C. Rubinsztein, M. Gill, B. Lawlor, A. Lynch, K. S. Brown, P. A. Passmore, D. Craig, B. McGuinness, S. Todd, C. Holmes, D. Mann, A. D. Smith, H. Beaumont, D. Warden, G. Wilcock, S. Love, P. G. Kehoe, N. M. Hooper, E. R. L. C. Vardy, J. Hardy, S. Mead, N. C. Fox, M. Rossor, J. Collinge, W. Maier, F. Jessen, E. Rüther, B. Schürmann, R. Heun, H. Kölsch, H. van den Bussche, I. Heuser, J. Kornhuber, J. Wiltfang, M. Dichgans, L. Frölich, H. Hampel, J. Gallacher, M. Hüll, D. Rujescu, I. Giegling, A. M. Goate, J. S. K. Kauwe, C. Cruchaga, P. Nowotny, J. C. Morris, K. Mayo, K. Sleegers, K. Bettens, S. Engelborghs, P. P. De Deyn, C. Van Broeckhoven, G. Livingston, N. J. Bass, H. Gurling, A. McQuillin, R. Gwilliam, P. Deloukas, A. Al-Chalabi, C. E. Shaw, M. Tsolaki, A. B. Singleton, R. Guerreiro, T. W. Mühleisen, M. M. Nöthen, S. Moebus, K.-H. Jöckel, N. Klopp, H.-E. Wichmann, V. S. Pankratz, S. B. Sando, J. O. Aasly, M. Barcikowska, Z. K. Wszolek, D. W. Dickson, N. R. Graff-Radford, R. C. Petersen, C. M. van Duijn, M. M. B. Breteler, M. A. Ikram, A. L. DeStefano, A. L. Fitzpatrick, O. Lopez, L. J. Launer, S. Seshadri, C. Berr, D. Campion, J. Epelbaum, J.-F. Dartigues, C. Tzourio, A. Alpérovitch, M. Lathrop, T. M. Feulner, P. Friedrich, C. Riehle, M. Krawczak, S. Schreiber, M. Mayhaus, S. Nicolhaus, S. Wagenpfeil, S. Steinberg, H. Stefansson, K. Stefansson, J. Snaedal, S. Björnsson, P. V Jonsson, V. Chouraki, B. Genier-Boley, M. Hiltunen, H. Soininen, O. Combarros, D. Zelenika, M. Delepine, M. J. Bullido, F. Pasquier, I. Mateo, A. Frank-Garcia, E. Porcellini, O. Hanon, E. Coto, V. Alvarez, P. Bosco, G. Siciliano, M. Mancuso, F. Panza, V. Solfrizzi, B. Nacmias, S. Sorbi, P. Bossù, P. Piccardi, B. Arosio, G. Annoni, D. Seripa, A. Pilotto, E. Scarpini, D. Galimberti, A. Brice, D. Hannequin, F. Licastro, L. Jones, P. A. Holmans, T. Jonsson, M. Riemenschneider, K. Morgan, S. G. Younkin, M. J. Owen, M. O’Donovan, P. Amouyel, and J. Williams, “Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer’s disease.,” *Nat. Genet.*, vol. 43, no. 5, pp. 429–35, 2011.

[7] M. J. Davies and D. A. Lomas, “The molecular aetiology of the serpinopathies,” *Int. J. Biochem. Cell Biol.*, vol. 40, no. 6–7, pp. 1273–1286, 2008.

[8] Q. Huang, L. Yang, J. Luo, L. Guo, Z. Wang, X. Yang, W. Jin, Y. Fang, J. Ye, B. Shan, and Y. Zhang, “SWATH enables precise label-free quantification on proteome scale,” *Proteomics*, vol. 15, no. 7, pp. 1215–1223, 2015.

[9] Y. Lyutvinskiy, H. Yang, D. Rutishauser, and R. A Zubarev, “In silico instrumental response correction improves precision of label-free proteomics and accuracy of proteomics-based predictive models.,” *Mol. Cell. Proteomics*, vol. 12, no. 8, pp. 2324–31, 2013.

[10] D. E. Clemmer, R. R. Hudgins, and M. F. Jarrold, “Naked protein conformations: cytochrome c in the gas phase,” *J. Am. Chem. Soc*, vol. 117, no. 40, pp. 10141–10142, 1995.

[11] K. Thalassinos, S. E. Slade, K. R. Jennings, J. H. Scrivens, K. Giles, J. Wildgoose, J. Hoyes, R. H. Bateman, and M. T. Bowers, “Ion mobility mass spectrometry of proteins in a modified commercial mass spectrometer,” *Int. J. Mass Spectrom*, vol. 236, no. 1–3, pp. 55–63, 2004.

[12] B. T. Ruotolo, J. L. P. Benesch, A. M. Sandercock, S.-J. Hyung, and C. V Robinson, “Ion mobility-mass spectrometry analysis of large protein complexes,” *Nat. Protoc.*, vol. 3, no. 7, pp. 1139–52, 2008.

[13] D. M. Walther, P. Kasturi, M. Zheng, S. Pinkert, G. Vecchi, P. Ciryam, R. I. Morimoto, C. M. Dobson, M. Vendruscolo, M. Mann, and F. U. Hartl, “Widespread Proteome Remodeling and Aggregation in Aging C. elegans,” *Cell*, vol. 161, no. 4, pp. 919–932, 2015.

[14] G. Favrin, D. M. Bean, E. Bilsland, H. Boyer, B. E. Fischer, S. Russell, D. C. Crowther, H. A. Baylis, S. G. Oliver, and M. E. Giannakou, “Identification of novel modifiers of Aβ toxicity by transcriptomic analysis in the fruitfly,” *Sci. Rep.*, vol. 3, p. 3512, 2013.

[15] M.-S. Kim, S. M. Pinto, D. Getnet, R. S. Nirujogi, S. S. Manda, R. Chaerkady, A. K. Madugundu, D. S. Kelkar, R. Isserlin, S. Jain, J. K. Thomas, B. Muthusamy, P. Leal-Rojas, P. Kumar, N. A. Sahasrabuddhe, L. Balakrishnan, J. Advani, B. George, S. Renuse, L. D. N. Selvan, A. H. Patil, V. Nanjappa, A. Radhakrishnan, S. Prasad, T. Subbannayya, R. Raju, M. Kumar, S. K. Sreenivasamurthy, A. Marimuthu, G. J. Sathe, S. Chavan, K. K. Datta, Y. Subbannayya, A. Sahu, S. D. Yelamanchi, S. Jayaram, P. Rajagopalan, J. Sharma, K. R. Murthy, N. Syed, R. Goel, A. A. Khan, S. Ahmad, G. Dey, K. Mudgal, A. Chatterjee, T.-C. Huang, J. Zhong, X. Wu, P. G. Shaw, D. Freed, M. S. Zahari, K. K. Mukherjee, S. Shankar, A. Mahadevan, H. Lam, C. J. Mitchell, S. K. Shankar, P. Satishchandra, J. T. Schroeder, R. Sirdeshmukh, A. Maitra, S. D. Leach, C. G. Drake, M. K. Halushka, T. S. K. Prasad, R. H. Hruban, C. L. Kerr, G. D. Bader, C. A. Iacobuzio-Donahue, H. Gowda, and A. Pandey, “A draft map of the human proteome,” *Nature*, vol. 509, no. 7502, pp. 575–81,2014.

[16] A. S. Hebert, A. L. Richards, D. J. Bailey, A. Ulbrich, E. E. Coughlin, M. S. Westphall, and J. J. Coon, “The One Hour Yeast Proteome,” *Mol. Cell. Proteomics*, vol. 13, pp. 339–347, 2013.

[17] L. C. Gillet, P. Navarro, S. Tate, H. Rost, N. Selevsek, L. Reiter, R. Bonner, and R. Aebersold, “Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis,” *Mol. Cell. Proteomics*, vol. 11, no. 6, pp. O111.016717–O111.016717, 2012.

[18] J. C. Silva, R. Denny, C. A. Dorschel, M. Gorenstein, I. J. Kass, G.-Z. Li, T. McKenna, M. J. Nold, K. Richardson, P. Young, and S. Geromanos, “Quantitative proteomic analysis by accurate mass retention time pairs,” *Anal. Chem.*, vol. 77, no. 7, pp. 2187–200,2005.

[19] U. Distler, J. Kuharev, P. Navarro, Y. Levin, H. Schild, and S. Tenzer, “Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics,” *Nat. Methods*, vol. 11, no. 2, pp. 167–70, 2014.

[20] R. Aebersold, A. L. Burlingame, and R. A. Bradshaw, “Western Blots vs. SRM Assays: Time to turn the tables?,” *Mol. Cell. Proteomics*, vol. 12, no. 9, pp. 2381–2382, 2013.

[21] R. A Zubarev, “The challenge of the proteome dynamic range and its implications for in-depth proteomics,” *Proteomics*, vol. 13, no. 5, pp. 723–6, 2013.

[22] J. R. Wiśniewski, M. Y. Hein, J. Cox, and M. Mann, “A ‘ proteomic ruler ’ for protein copy number and concentration estimation without spike - in standards,” *Mol. Cell. Proteomics*, vol. 13, no. 12, pp. 3497–3506, 2014.

[23] J. C. Silva, M. V Gorenstein, G.-Z. Li, J. P. C. Vissers, and S. J. Geromanos, “Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition.,” *Mol. Cell. Proteomics*, vol. 5, pp. 144–156, 2006.

[24] J. M. Chick, D. Kolippakkam, D. P. Nusinow, B. Zhai, R. Rad, E. L. Huttlin, and S. P. Gygi, “A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides,” *Nat. Biotechnol.*, vol. 33, no. 7, pp. 743–749, 2015.

APPENDIX A

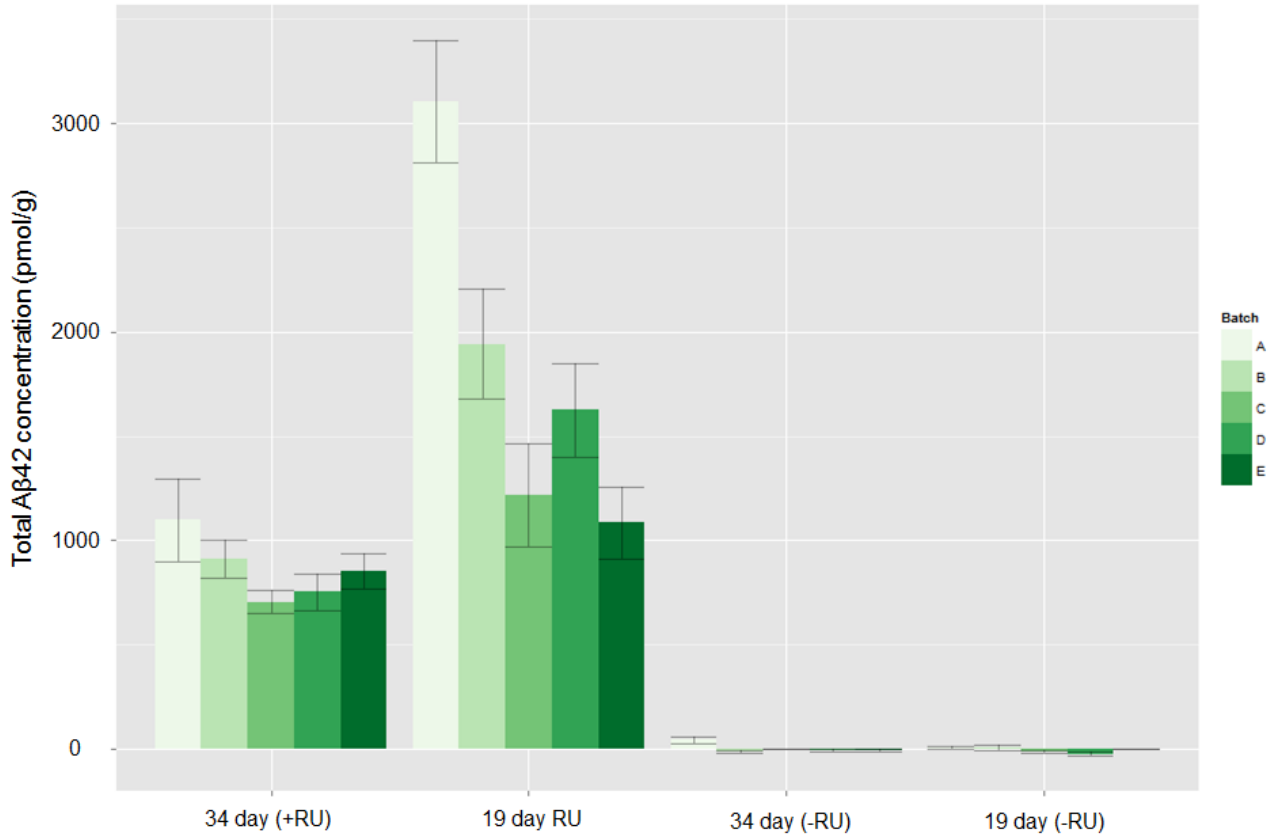
R script for LOESS peptide level normalisation

```
MSdata <- read.csv('test data for MA3.csv')        # read in peptide intensity matrix

MSdata[MSdata == 0] <- NA                          # replace 0 values with NA

cor <- cor(MSdata, use= 'complete')

coravg <- colMeans(cor, na.rm=T)                   # calculate average correlation

coravg1 <- as.matrix(coravg)

ref <- which.max(coravg1[,1])                      # record which  column to which all other are
normalised


M <- matrix(nrow=nrow(MSdata), ncol=(ncol(MSdata))) {

        for (i in 1:ncol(M)) {M[,i] = log2(MSdata[,ref]/MSdata[,i])          # calculate M values

}

A <- matrix(nrow=nrow(MSdata),ncol=ncol(MSdata)) {

        for (i in 1:ncol(A)) {A[,i] = log2(0.5*(MSdata[,ref]*MSdata[,i]))    # calculate A values

}


Mcor <- matrix(nrow=nrow(M),ncol=ncol(M))                                    # loess regression

for(i in 1:ncol(M)){

        l <- loess(M[,i]~A[,i])

        pred <- predict(l,A[,i])

        Mcor[,i] <- pred

}


Mnew <- matrix(nrow=nrow(M),ncol=ncol(M))                  # calculate normalised M values

for (i in 1:ncol(M)) {

        Mnew[,i] <- M[,i]-Mcor[,i]
```

```
}                                                    # back calculates normalised intensities

MSnormintensity <- matrix(nrow=nrow(M),ncol=ncol(M))


for(i in 1:ncol(M)) {

        MSnormintensity[,i] <- PeptideIntensity[,1]/2^(Mnew[,i])

}
```

Appendix B

Quantification of total Aβ42 in four different conditions and five different batches of an inducible *D. melanogaster* AD model (Chapter 2)



Appendix C

Tabulated results from binary comparison (Chapter 2)

Old + / Old -

Accession	Protein	Function	Fold (34P/34M)	Effect Size (Cohen's d)	P value
Q9W3L4	CG2233	Unknown	0.49	1.98	0.00039
P29844	Heat shock 70Kda protein cognate 3	Assembly of multimeric protein complexs inside ER	1.47	1.13	0.00069
Q7JWQ7	CG3074	Immune response and proteolysis.Belongs to peptidase C1 family	2.24	2.16	0.00287
P25228	Ras-related protein Rab-3	Exocytosis. Neurotransmitter release	0.6	1.49	0.0043
Q961R8	Glycyl-tRNA synthase	Aminoacyl-tRNA synthase	1.56	1.3	0.00559
P22700-2	Calcium-transporting ATPase sarcoplasmic/endoplasmic reticulum type	Catalysis of ATO hydrolysis	1.84	1.72	0.00768

Old - / Young –

Accession	**Protein**	**Function**	**Fold (34d-/19d-)**	**Effect Size (Cohen's d)**	**P value**
P33438	Glutactin	Calcium Binding	1.49	1.18	0.001475774
Q9W3L4	CG2233	Unknown	2.30	2.22	0.00247914
B7Z061	Photoreceptor dehydrogenase	Oxidoreductase	0.59	1.51	0.003687654
P40417-3	Mitogen-activated protein kinase ERK-A	Signal transduction pathways	0.55	1.68	0.003741004
Q9U915	Adenylate kinase	Cellular energy homeostasis and metabolism	0.70	1.06	0.003807227
Q0E9G4	CG1600	Oxidoreductase	0.55	1.67	0.006552275

Old + / Young +

Accession	Protein	Function	Fold (34d+/19d+)	Effect Size (Cohen's d)	P value
Q961R8	Glycyl-tRNA synthetase	Aminoacyl-tRNA synthetase	1.53	1.25	0.0002925
Q95028	L-lactate dehydrogenase	Glycolysis	0.47	2.07	0.007475281
O62530	Clathrin-associated protein AP2 (AP50)	Protein transport. Neurotransmitter secretion	0.61	1.42	0.009298651
Q7K084	RH04549p	Sensory perception of chemical stimulus	0.68	1.11	0.010200116
Q960M4	Peroxiredoxin 5	Oxidoreductase	1.64	1.43	0.010789987

Young + / Young –

Accession	Protein	Function	Fold (19P/19M)	Effect Size (Cohen's d)	P value
P29844	Heat shock 70 kDa protein cognate 3	Chaperone	2.06	1.99	2.15091E-07
Q960M4	Peroxiredoxin 5	Oxidoreductase	0.62	1.40	0.001742624
Q9W5W8	CG9577	Isomerase	0.65	1.27	0.004163219
Q24583	V-type proton ATPase subunit F 1	Ion transport	1.40	1.00	0.006504416
P25455-2	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase classes I and II	Lipid degradation	1.98	1.90	0.009649766

Appendix D

Proteins defined as unique to a condition from pairwise comparisons

Old- unique (vs old+)

Accession	Protein	Number of Bios old +	Number of Technical old +	Number of Bios old -
P92177-4	14-3-3 protein epsilon	4	6	0
Q24086	Soluble guanylyl cyclase beta subunit	3	4	0
Q7JVK6	Translin	3	4	0
A4V4A5	Ran	3	3	0
E1JHJ4	Myosin heavy chain	3	3	0
Q08012	Protein enhancer of sevenless 2B	3	3	0
Q4QQA3	ProteinC kinase 98E	3	3	0
Q9VVU2	Ribosomal protein L26	3	3	0
O15971	Rab 10	3	3	0
Q9VM50	Rab 30	3	3	0
Q9W3K6	CG2258	3	3	0
A1ZB71	Glutathione S transferase E6	2	3	0
P11146	Heat shock 70 protein cognate 2	2	3	0
Q7K2D2	Probbale dynactin subunit 2 (Dynamitin)	2	3	0
Q7KN62-2	**Transitional endoplasmic reticulum ATPase TER94**	2	3	0
Q8INN5	Unc-115-Isoform B	2	3	0
Q8MT29	Tom 70	2	3	0
Q9VY05	GH11762p	2	3	0
O97471	Microsomal glutathione S-transferase-like protein	2	3	0

Young - unique (vs old +)

Protein	Number of Bios 19day-	Number of Technicals 19day-	Number of Bios 34dayM
14-3-3 protein epsilon	4	5	0
CG33303	3	5	0
Cytoplasmic phosphatidylinositol transfer protein 1	2	4	0
CG10077	3	4	0
Eukaryotic translation initiation factor 3 subunit 5-1	2	4	0
Putative peptidyl-prolyl cis-trans isomerase dodo	2	4	0
Cellular retinaldehyde binding protein	3	3	0
CG6613	2	3	0
ATPase ASNA1 homolog	2	3	0
Actin, larval muscle	2	3	0
26S protease regulatory subunit 4	2	3	0
Bitesize isoform 3 (granuphiliin homolog) and synaptotagmin-like protein	2	3	0
Mustard	2	3	0
CG6463	2	3	0
Acyl carrier protein, mitochondrial	2	3	0
60S ribosomal protein L31	2	3	0
Vitellogenin-2	2	3	0
Heat shock protein 27	2	3	0
CG31281	2	3	0
Cytochrome b-c1 complex subunit 9	2	3	0
F-actin-capping protein subunit beta	2	3	0
Eukaryotic translation initiation factor 5A	2	3	0
CG30378	2	3	0
CG8888	2	3	0

Young + unique (vs old +)

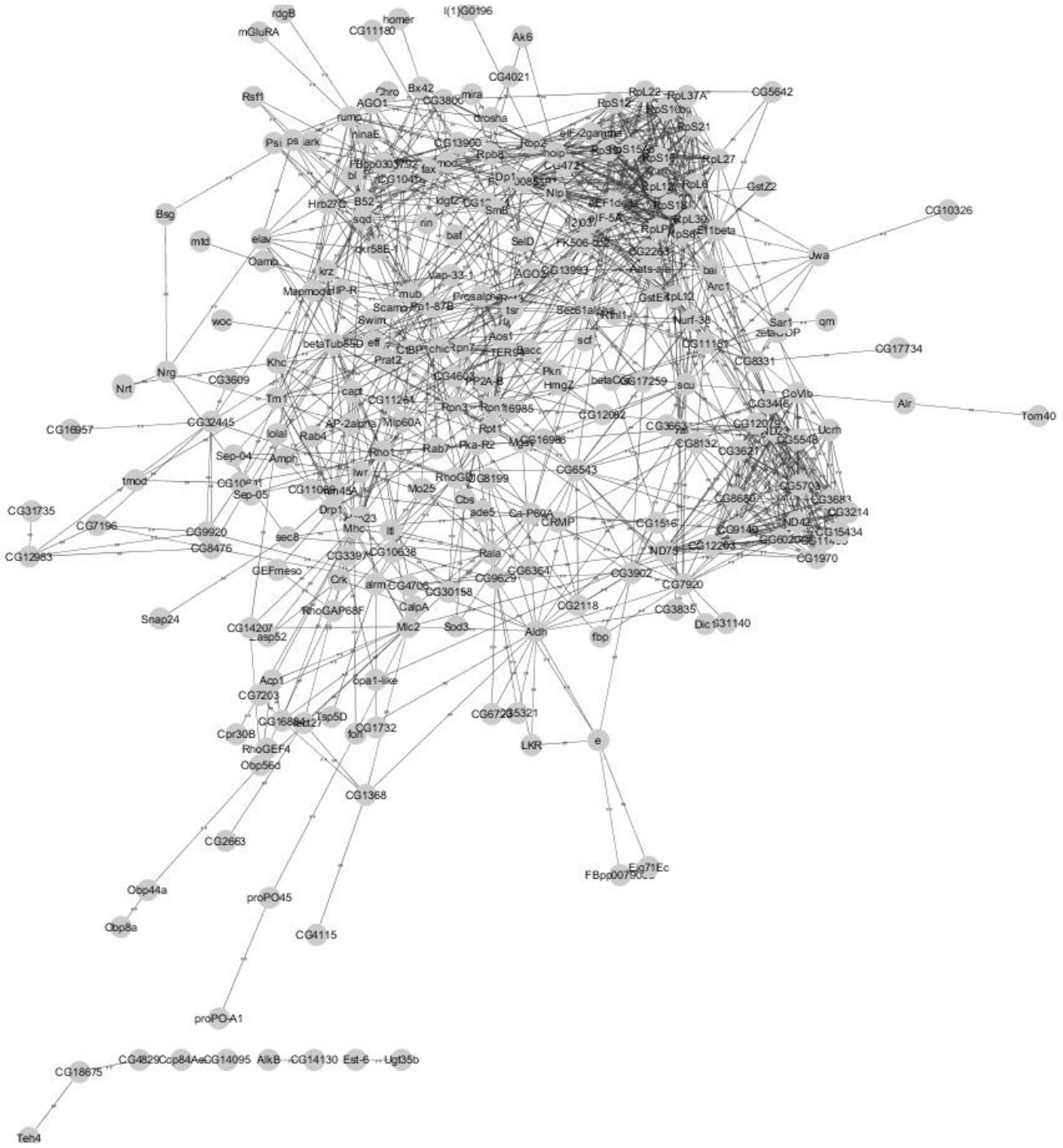
Protein	Number of Bios 34day-	Number of Technicals 34day-	Number of Bios 34day+
NAP1	4	6	0
14-3-3 protein epsilon	4	5	0
HSP27	4	5	0
Heterogeneous nuclear ribonucleoprotein A1	3	5	0
Microtubule-assocaited protein Jupiter	4	5	0
cAMP-dependent protein kinase type I	4	5	0
Eukaryotic translation initiation factor 3 subunit F-1	3	4	0
Heat shock protein 26	4	4	0
Heat shock 70 kDa protein cognate 2	2	4	0
Myosin regulatory light chain 2	3	4	0
CG9090	2	4	0
Proteasome subunit beta type	3	4	0
Eb1, isoform F	2	3	0
Sec22	2	3	0
Myofilin	2	3	0
Myosin heavy chain, muscle	2	3	0
Dihydroorotate dehydrogenase (quinone, mitochonrial	3	3	0
Flightin	2	3	0
Troponin1	2	3	0
Tetraspanin 42Ef	2	3	0
CG11474	2	3	0
Sarcoplasmic calcium-binding protein 1	2	3	0
IA-2	2	3	0
cg991	2	3	0
40s ribosomal protein s5a	2	3	0
cg9512	2	3	0
Glycogen Synthase	3	3	0

Young + unique (vs young -)

Protein	Number of Bios 19dayP	Number of Technicals 19dayP	Number of Bios 19dayM
Proteasome subunit alpha type-7-1	4	7	0
Heat shock protein 26	4	4	0
tRNA (guanine-N(7)-)-methyltransferase non-catalytic subunit wuho	3	5	0
Proteasome subunit beta type	3	4	0
Myosin regulatory light chain 2	3	4	0
Dihydroorotate dehydrogenase (quinone), mitochondrial	3	3	0
60S ribosomal protein L7a	2	4	0
Myosin heavy chain, muscle	2	3	0
Protein kinase shaggy (differences in isoforms)	2	3	0
Flightin	2	3	0
Putative hydroxypyruvate isomerase	2	3	0
Sarcoplasmic calcium-binding protein 1	2	3	0
CG9911	2	3	0
40S ribosomal protein S28	2	3	0

Appendix E

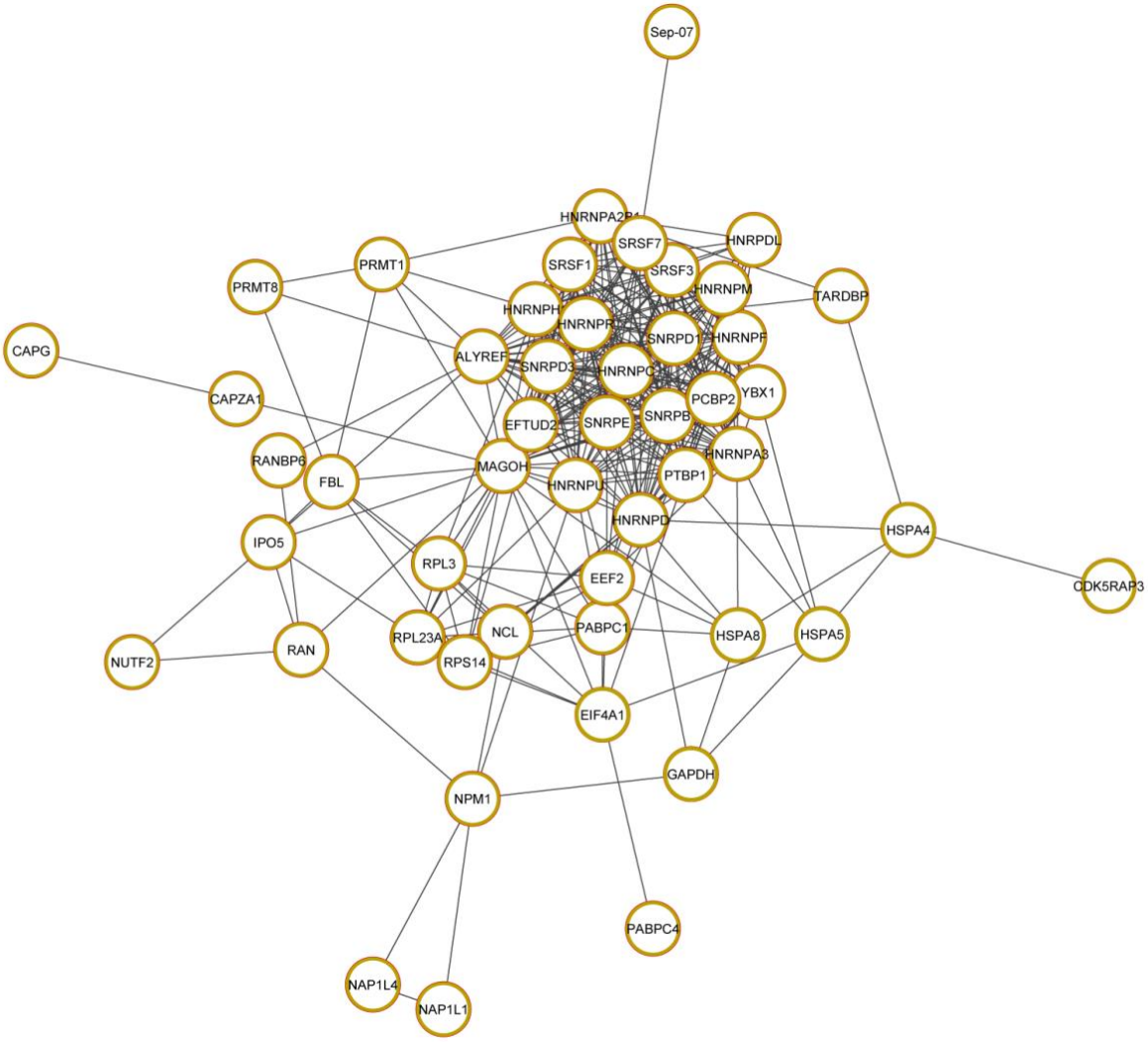
Ageing protein-protein interaction network



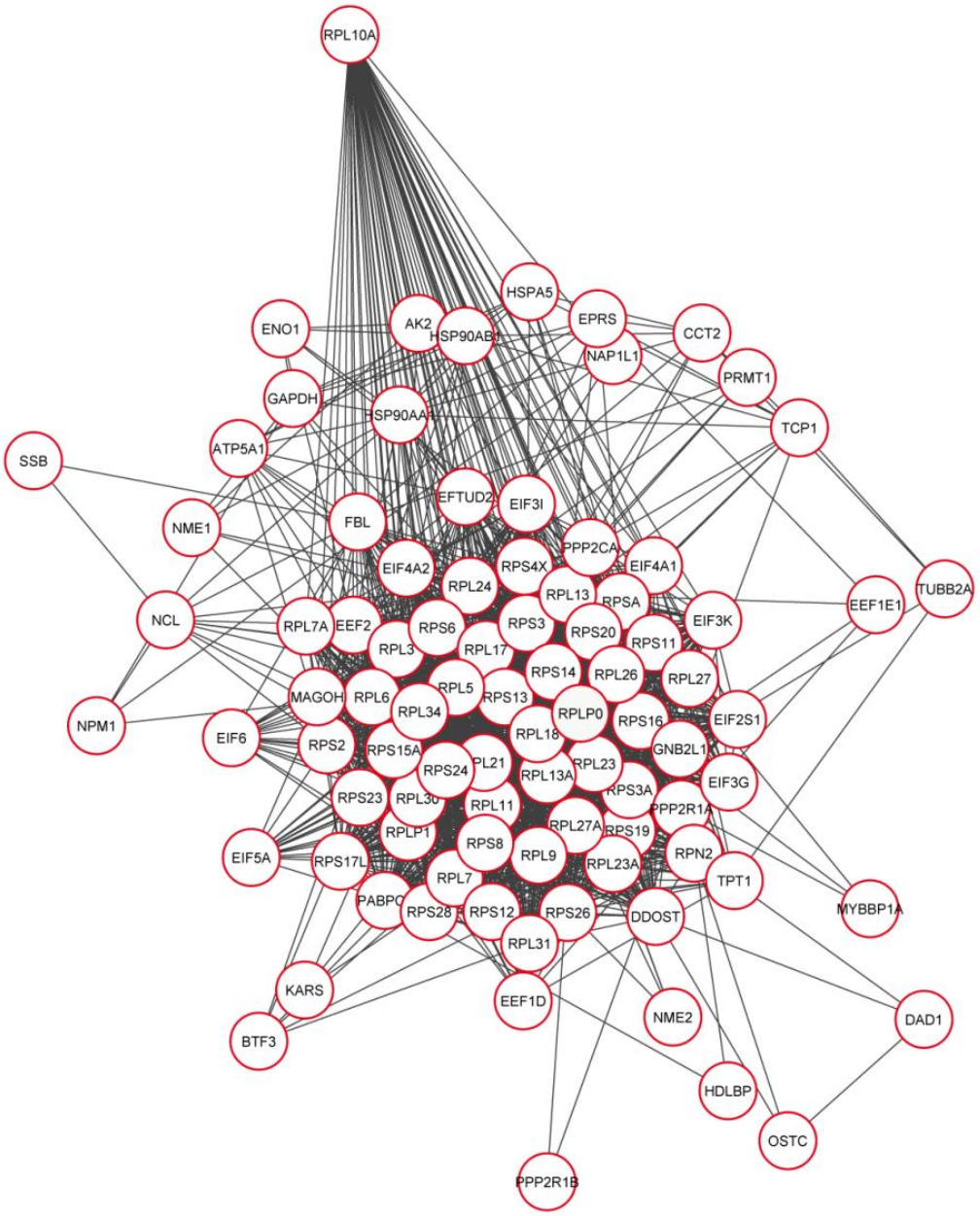
APPENDIX F

Protein-protein interaction networks for protein >1.3 fold upregulated, NHK+/M+ (Chapter 4)

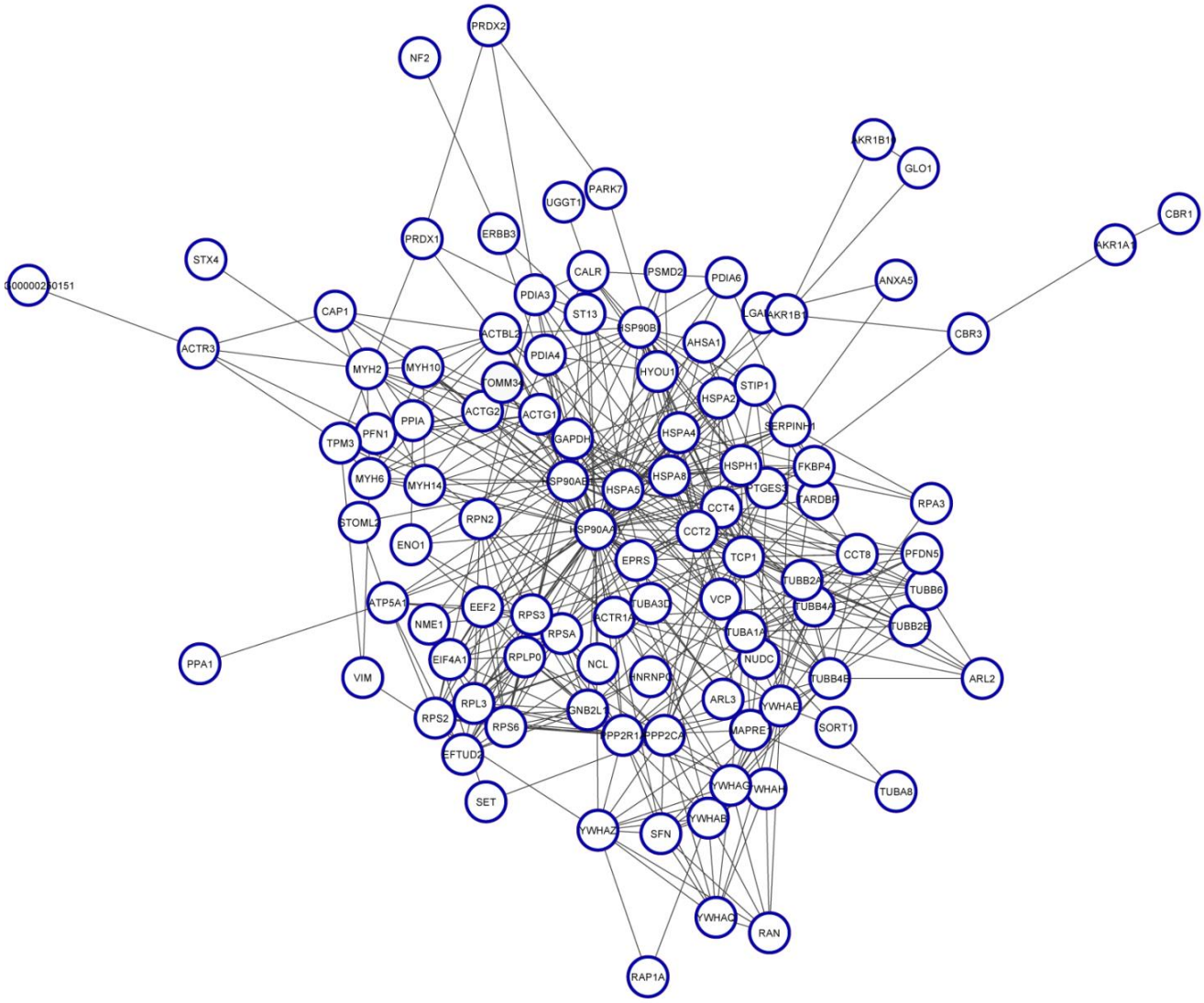
Cluster I



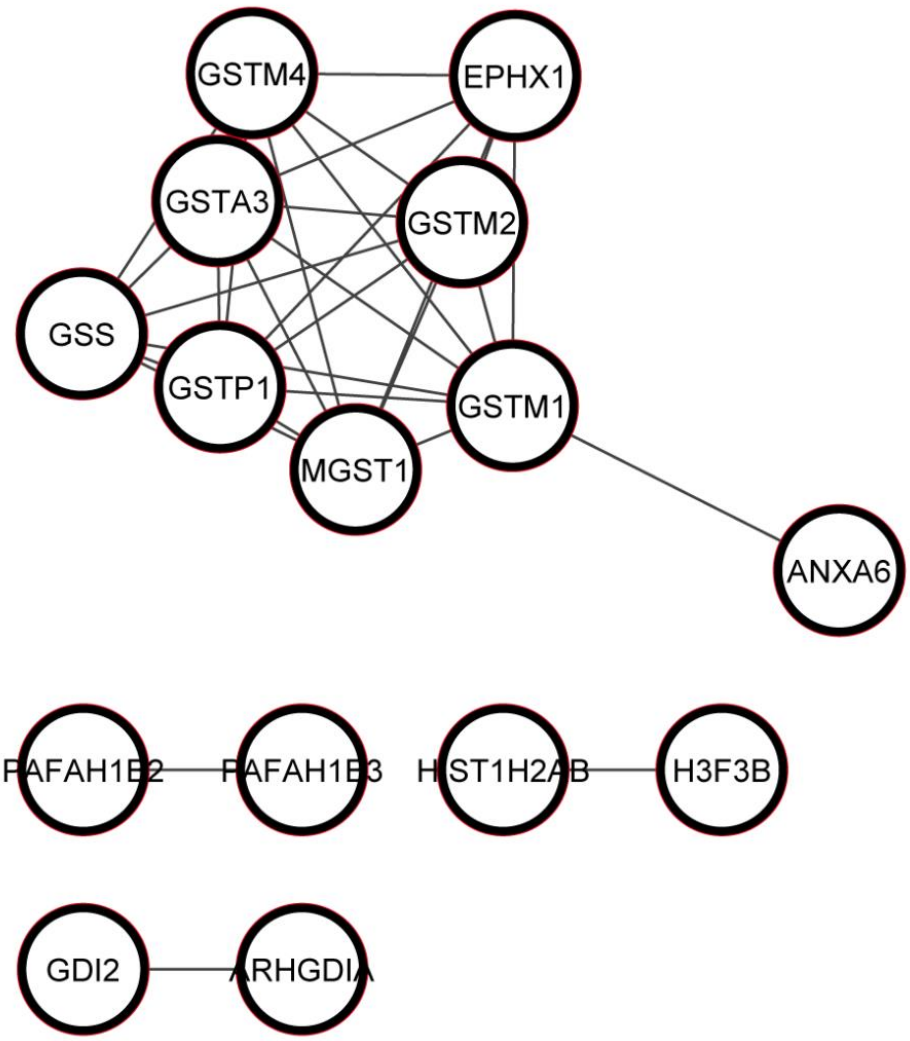
Cluster II


RPL10A
HSPA5
EPRS
ENO1
AK2
HSP90AB1
CCT2
NAP1L1
PRMT1
GAPDH
HSP90AA1
TCP1
ATP5A1
SSB
EFTUD2
EIF3I
NME1
FBL
PPP2CA
EIF4A2
RPS4X
EIF4A1
RPL24
RPL13
RPSA
TUBB2A
EEF1E1
NCL
RPL7A
EEF2
RPS6
RPS3
RPS20
EIF3K
RPL3
RPL17
RPS11
RPS14
RPL26
RPL27
MAGOH
RPL6
RPL5
RPS13
RPL34
RPLP0
RPS16
EIF2S1
NPM1
EIF6
RPS2
RPS15A
RPL18
GNB2L1
RPS24
RPL21
RPL13A
RPL23
EIF3G
RPS23
RPL30
RPL11
RPS3A
PPP2R1A
RPLP1
RPL27A
RPS19
EIF5A
RPS8
RPL9
RPN2
RPS17L
RPL23A
TPT1
RPL7
PABPC
RPS28
RPS12
RPS26
DDOST
MYBBP1A
RPL31
KARS
EEF1D
NME2
DAD1
BTF3
HDLBP
OSTC
PPP2R1B

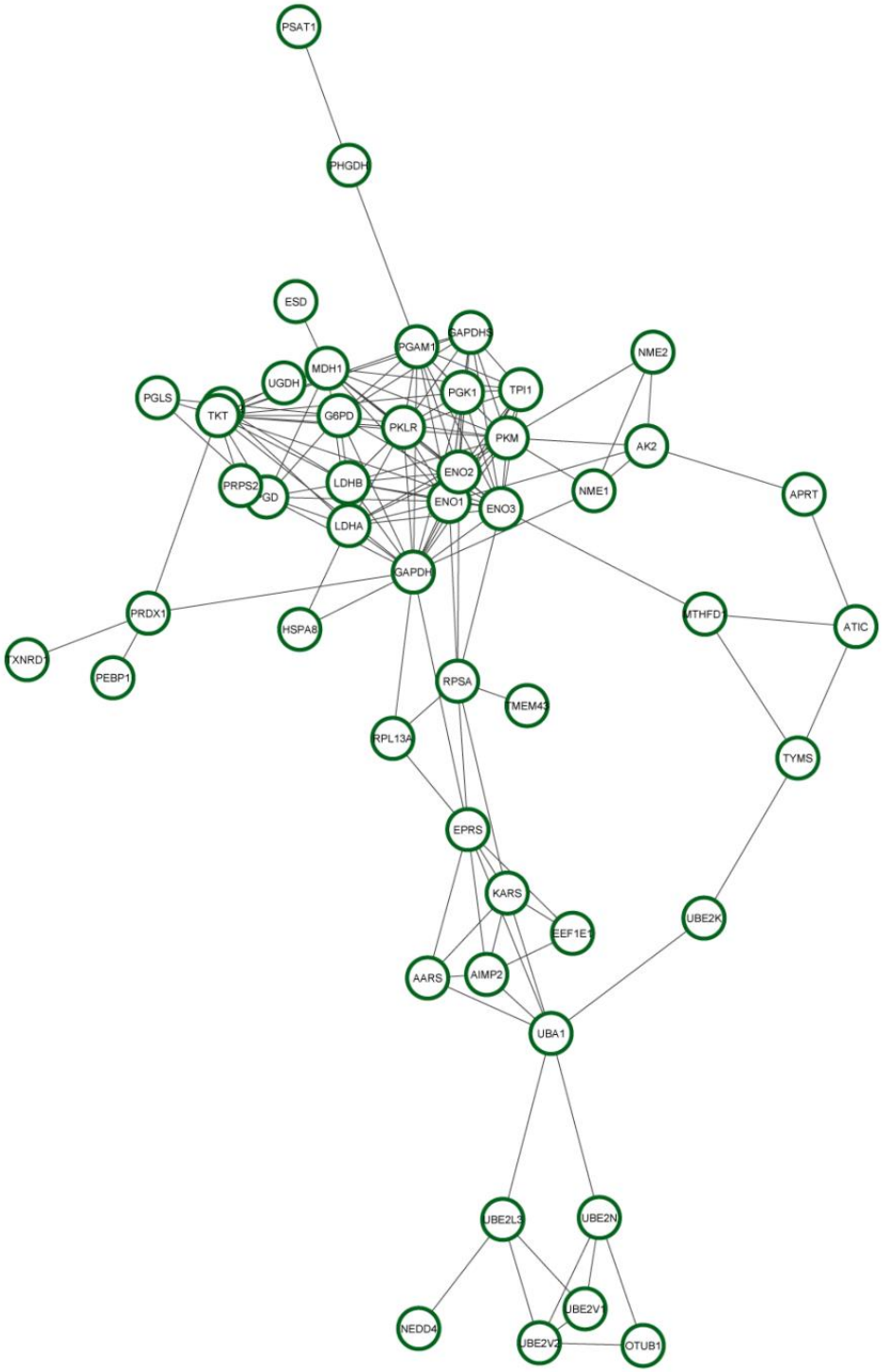
Cluster III


PRDX2
NF2
AKR1B1
GLO1
UGGT1
PARK7
CBR1
AKR1A1
PRDX1
ERBB3
STX4
CALR
PSMD2
PDIA6
ANXA5
ENSG00000230151
PDIA3
ST13
HSP90B1
LGALS1
AKR1B10
CAP1
ACTR3
ACTBL2
PDIA4
AHSA1
CBR3
MYH2
MYH10
TOMM34
HYOU1
STIP1
HSPA2
ACTG2
ACTG1
PPIA
PFN1
HSPA4
SERPINH1
TPM3
GAPDH
HSPH1
FKBP4
MYH6
MYH14
HSP90AB1
HSPA8
HSPA5
PTGES3
TARDBP
RPA3
STOML2
RPN2
CCT4
HSP90AA1
CCT2
ENO1
EPRS
TCP1
CCT8
PFDN5
TUBB2A
TUBB6
ATP5A1
EEF2
RPS3
TUBA3D
VCP
TUBB4A
NME1
ACTR1A
RPSA
TUBA1A
TUBB2B
PPA1
EIF4A1
RPLP0
NCL
NUDC
VIM
HNRNPC
TUBB4B
ARL2
RPL3
ARL3
YWHAE
RPS2
RPS6
SNB2L1
PPP2R1A
PPP2CA
MAPRE1
SORT1
EFTUD2
YWHAG
YWHAH
TUBA8
SET
YWHAB
YWHAZ
SFN
YWHAQ
RAN
RAP1A

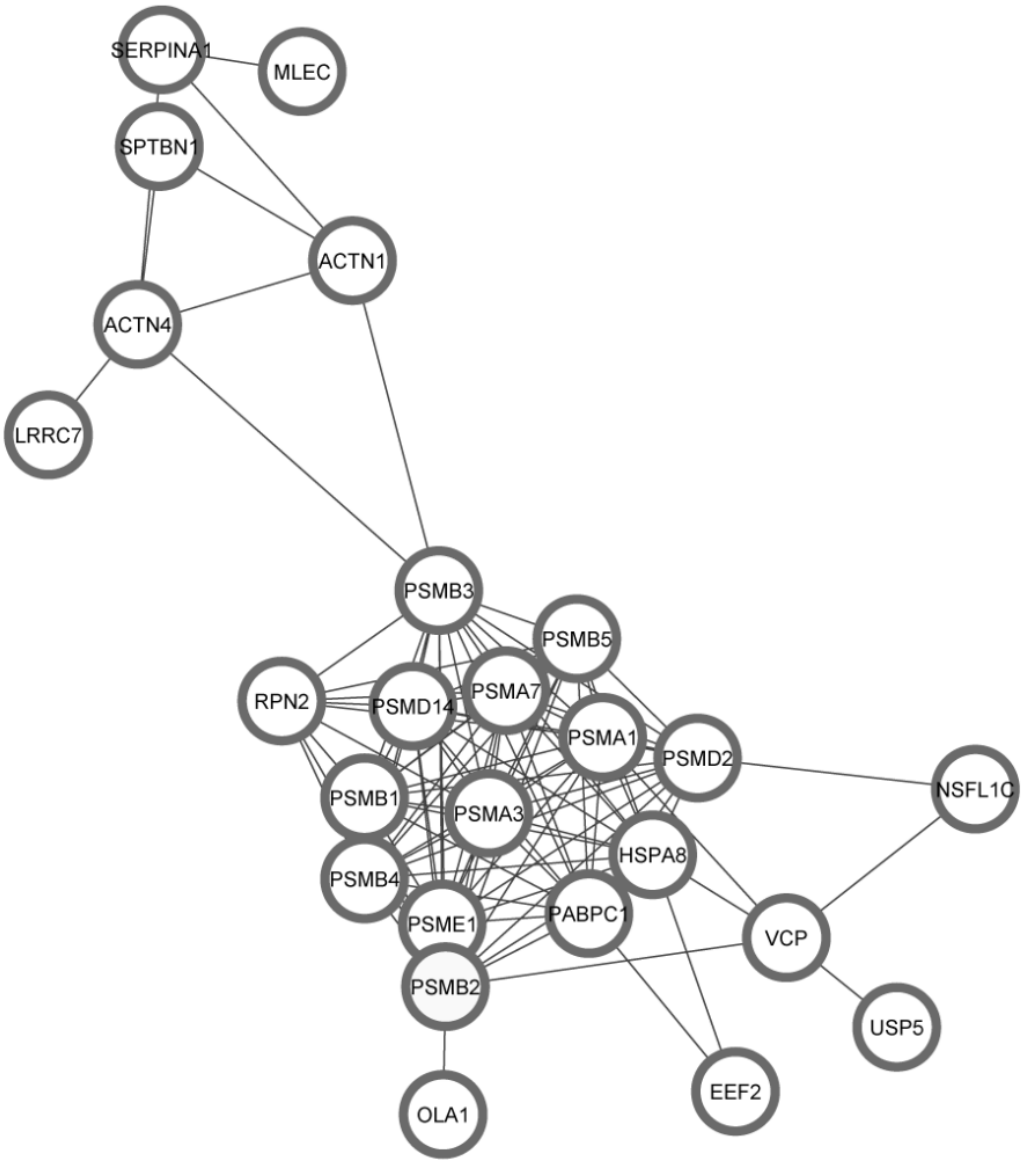
Cluster IV


GSTM4
EPHX1
GSTA3
GSTM2
GSS
GSTP1
GSTM1
MGST1
ANXA6
PAFAH1B2
PAFAH1B3
HIST1H2AB
H3F3B
GDI2
ARHGDIA

Cluster V

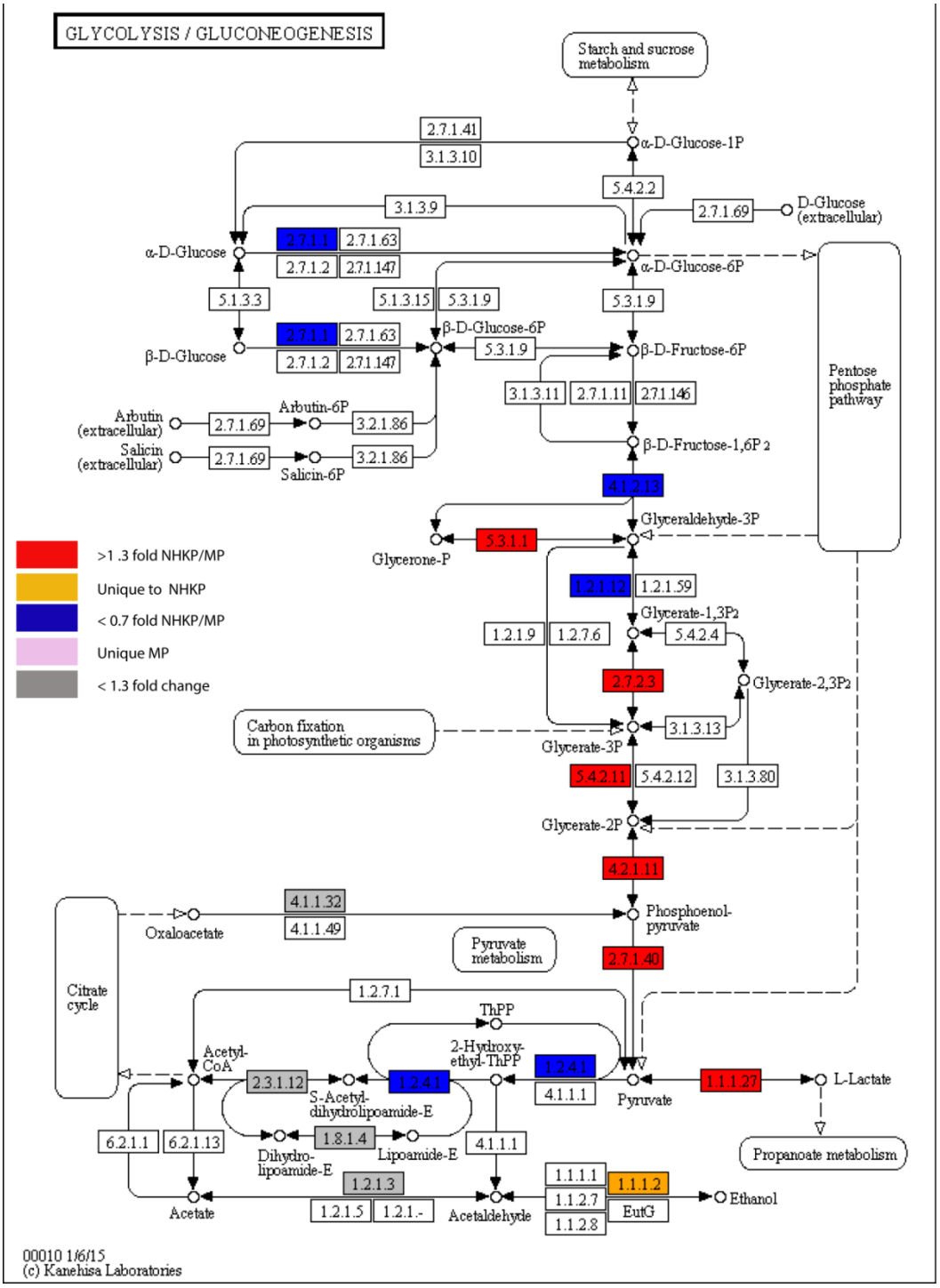

PSAT1
PHGDH
ESD
GAPDHS
PGAM1
NME2
MDH1
UGDH
PGLS
TKT
PGK1
TPI1
G6PD
PKLR
PKM
AK2
PRPS2
GD
LDHB
ENO2
NME1
APRT
ENO1
ENO3
LDHA
GAPDH
PRDX1
MTHFD
ATIC
HSPA8
TXNRD
PEBP1
RPSA
MEM4
RPL13A
TYMS
EPRS
KARS
UBE2K
EEF1E
AARS
AIMP2
UBA1
UBE2L3
UBE2N
UBE2V
NEDD4
UBE2V2
OTUB1

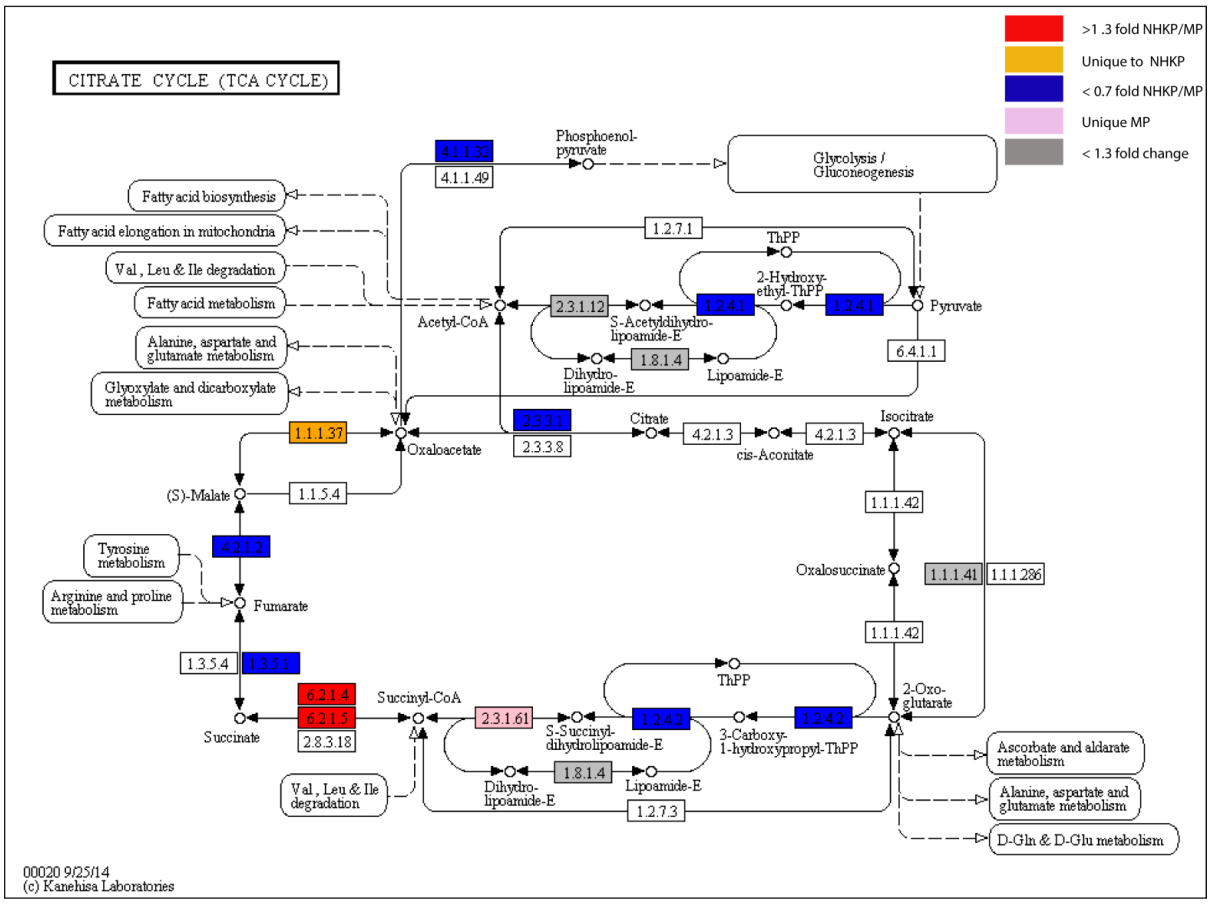
Cluster VI

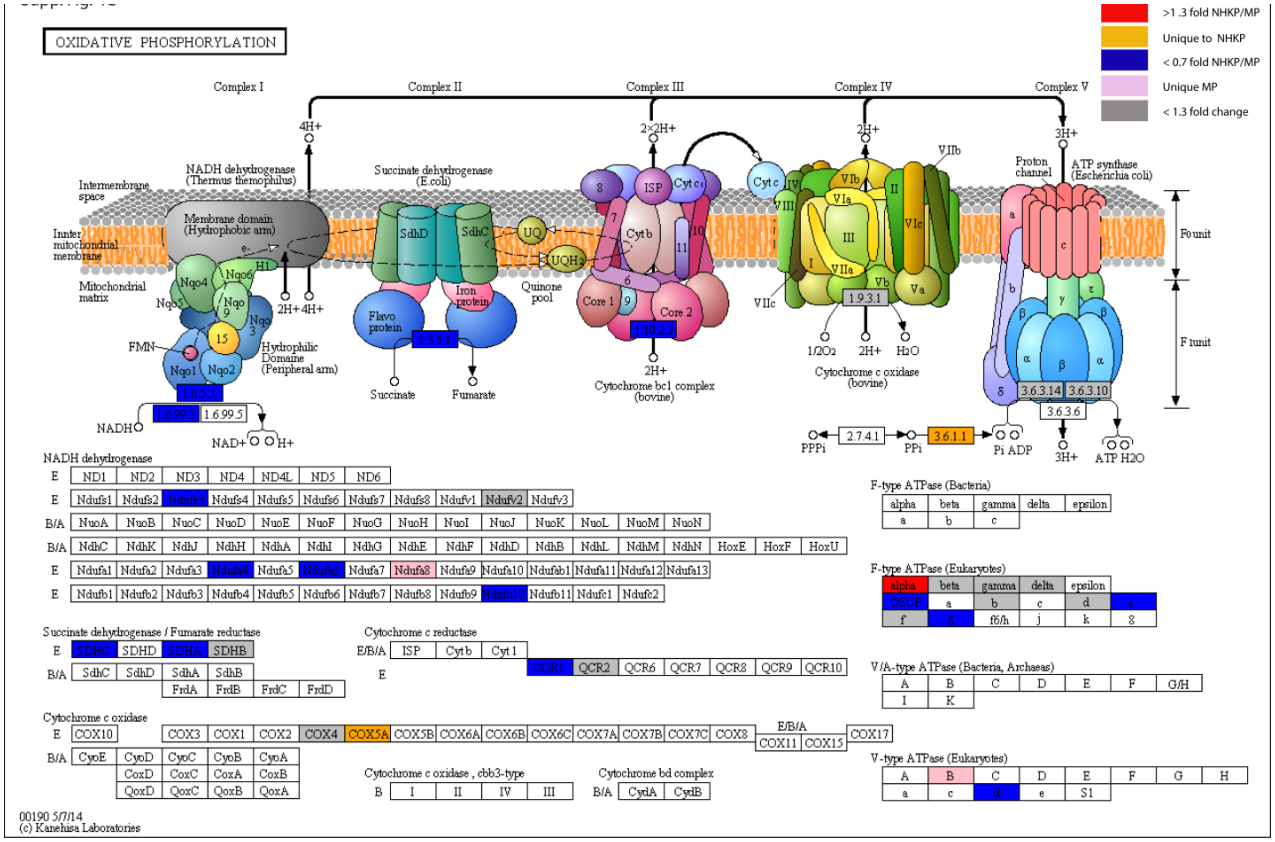

SERPINA1
MLEC
SPTBN1
ACTN1
ACTN4
LRRC7
PSMB3
PSMB5
RPN2
PSMD14
PSMA7
PSMA1
PSMD2
NSFL1C
PSMB1
PSMA3
HSPA8
PSMB4
PSME1
PABPC1
VCP
PSMB2
USP5
EEF2
OLA1

APPENDIX G

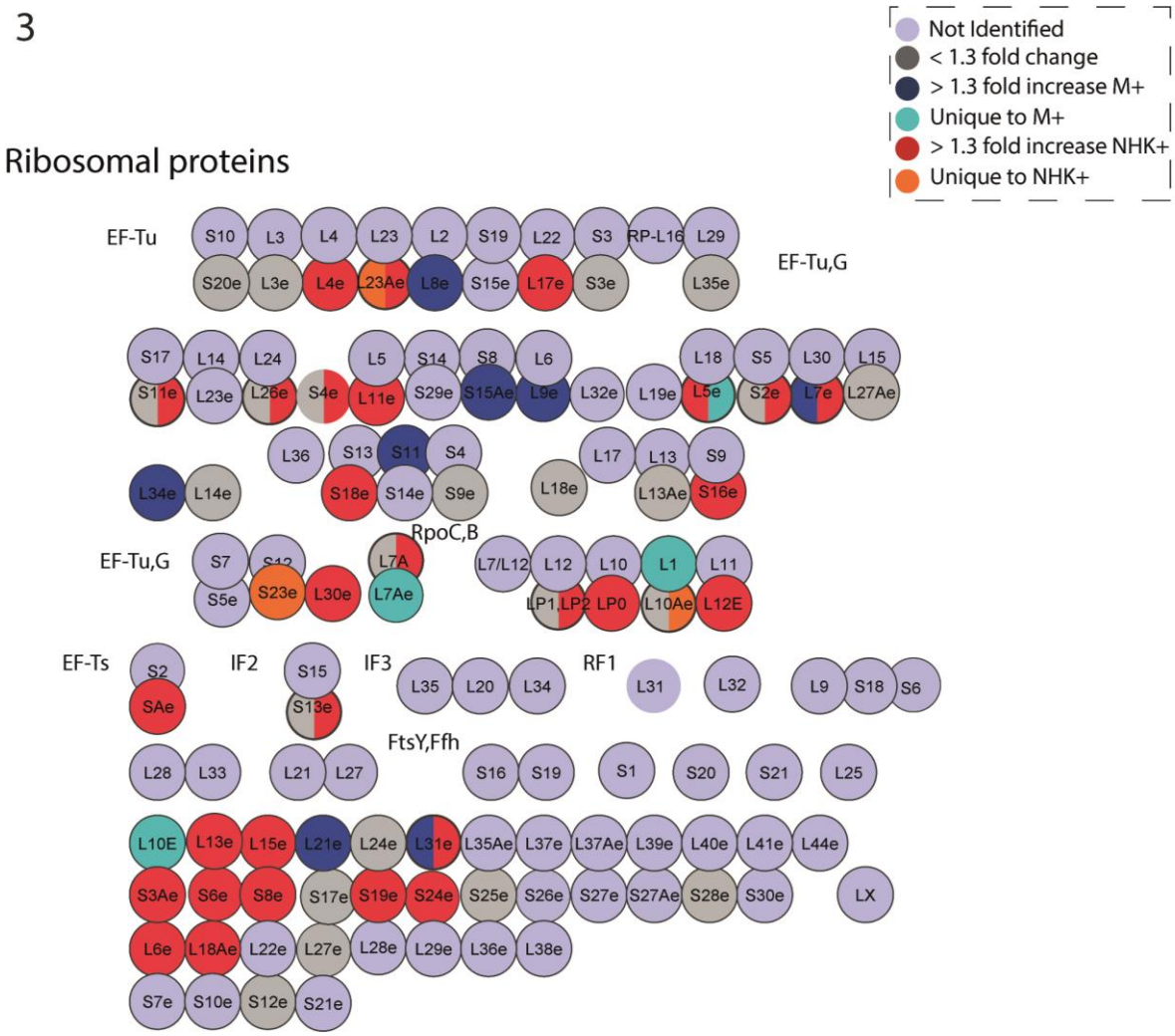
KEGG pathway analysis, NHK+ vs M+ (Chapter 4)




CITRATE CYCLE (TCA CYCLE)
>1 .3 fold NHKP/MP
Unique to NHKP
< 0.7 fold NHKP/MP
Unique MP
< 1.3 fold change
Phosphoenol-pyruvate
4.1.1.32
4.1.1.49
Glycolysis / Gluconeogenesis
Fatty acid biosynthesis
Fatty acid elongation in mitochondria
Val , Leu & Ile degradation
Fatty acid metabolism
Alanine, aspartate and glutamate metabolism
Glyoxylate and dicarboxylate metabolism
1.2.7.1
ThPP
2-Hydroxy-ethyl-ThPP
Acetyl-CoA
2.3.1.12
S-Acetyldihydro-lipoamide-E
1.2.4.1
1.2.4.1
Pyruvate
Dihydro-lipoamide-E
1.8.1.4
Lipoamide-E
6.4.1.1
1.1.1.37
2.3.3.1
2.3.3.8
Oxaloacetate
Citrate
4.2.1.3
4.2.1.3
cis-Aconitate
Isocitrate
(S)-Malate
1.1.5.4
1.1.1.42
Tyrosine metabolism
4.2.1.2
Arginine and proline metabolism
Fumarate
Oxalosuccinate
1.1.1.41
1.1.1.286
1.1.1.42
1.3.5.4
1.3.5.1
ThPP
6.2.1.4
6.2.1.5
2.8.3.18
Succinyl-CoA
Succinate
2.3.1.61
S-Succinyl-dihydrolipoamide-E
1.2.4.2
1.2.4.2
2-Oxo-glutarate
3-Carboxy-1-hydroxypropyl-ThPP
Ascorbate and aldarate metabolism
Alanine, aspartate and glutamate metabolism
D-Gln & D-Glu metabolism
Val , Leu & Ile degradation
Dihydro-lipoamide-E
1.8.1.4
Lipoamide-E
1.2.7.3
00020 9/25/14
(c) Kanehisa Laboratories


OXIDATIVE PHOSPHORYLATION
>1.3 fold NHKP/MP
Unique to NHKP
< 0.7 fold NHKP/MP
Unique MP
< 1.3 fold change
Complex I
Complex II
Complex III
Complex IV
Complex V
NADH dehydrogenase (Thermus thermophilus)
Succinate dehydrogenase (E.coli)
Proton channel
ATP synthase (Escherichia coli)
Intermembrane space
Membrane domain (Hydrophobic arm)
Inner mitochondrial membrane
Mitochondrial matrix
Hydrophilic Domain (Peripheral arm)
Flavo protein
Iron protein
Quinone pool
Succinate
Fumarate
Cytochrome bc1 complex (bovine)
Cytochrome c oxidase (bovine)
F0unit
F1unit
NADH
NAD+
FMN
PPPi
PPi
Pi ADP
ATP H2O
NADH dehydrogenase
Succinate dehydrogenase / Fumarate reductase
Cytochrome c reductase
Cytochrome c oxidase
Cytochrome c oxidase , cbb3-type
Cytochrome bd complex
F-type ATPase (Bacteria)
F-type ATPase (Eukaryotes)
V/A-type ATPase (Bacteria, Archaeas)
V-type ATPase (Eukaryotes)
00190 5/7/14
(c) Kanehisa Laboratories

Appendix H


3
Not Identified
< 1.3 fold change
> 1.3 fold increase M+
Unique to M+
> 1.3 fold increase NHK+
Unique to NHK+
Ribosomal proteins
EF-Tu
EF-Tu,G
RpoC,B
EF-Tu,G
EF-Ts
IF2
IF3
RF1
FtsY,Ffh